GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 22, 2002, 12:47:48 ; Search time 2161.72 Seconds
(without alignments)
193.610 Million cell updates/sec

Title: US-09-747-514A-1
Perfect score: 20
Sequence: 1 ccgcgctcggccgcagcctg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	20	100.0	3123	1	PAU38241	U38241 Pseudomonas
c 2	20	100.0	13075	1	AE004946	AE004946 Pseudomon
3	18.4	92.0	43903	2	AC004396	AC004396 Pseudomon
c 4	17.4	87.0	650	9	HSA335678	AJ335678 Homo sapi
c 5	17.4	87.0	690	9	HSA337098	AJ337098 Homo sapi
6	17.4	87.0	970	9	HS2NRG12	AF119162 Homo sapi
7	17.4	87.0	1889	8	AB063178	AB063178 Hordeum v
8	17.4	87.0	2268	6	AR098155	AR098155 Sequence
9	17.4	87.0	2268	6	AR116627	AR116627 Sequence
10	17.4	87.0	3020	9	AB005060	AB005060 Homo sapi
c 11	17.4	87.0	3119	1	DVU16723	U16723 Desulfovibr
12	17.4	87.0	3777	8	OSA427974	AJ427974 Oryza sat
c 13	17.4	87.0	4160	6	A06936	A06936 H.sapiens f
c 14	17.4	87.0	4160	6	E01406	E01406 DNA sequenc
c 15	17.4	87.0	7335	6	AX097467	AX097467 Sequence
c 16	17.4	87.0	9180	1	AF274704	AF274704 Pseudomon
c 17	17.4	87.0	12263	9	HSFESFPS	X06292 Human c-fes
18	17.4	87.0	15719	1	AE004667	AE004667 Pseudomon
19	17.4	87.0	17200	1	AF293442	AF293442 Streptomy
c 20	17.4	87.0	18023	1	AX024392	AX024392 Sequence
c 21	17.4	87.0	18023	6	AX024285	AX024285 Sequence
22	17.4	87.0	28804	1	SSU51197	U51197 Sphingomona
23	17.4	87.0	28804	6	AR068625	AR068625 Sequence
c 24	17.4	87.0	105198	8	AC011806	AC011806 Genomic S
c 25	17.4	87.0	110000	2	LMFLCHR34_09	Continuation (10 o
c 26	17.4	87.0	145223	2	AF289077	AF289077 Homo sapi
c 27	17.4	87.0	153015	2	AP004463	AP004463 Oryza sat
c 28	17.4	87.0	153197	9	AC023306	AC023306 Homo sapi
c 29	17.4	87.0	154716	2	AC069502	AC069502 Homo sapi
c 30	17.4	87.0	156089	2	AC004586	AC004586 Homo sapi
c 31	17.4	87.0	157678	9	AC084222	AC084222 Homo sapi
c 32	17.4	87.0	161720	8	AP003281	AP003281 Oryza sat
c 33	17.4	87.0	167399	8	AP003236	AP003236 Oryza sat
c 34	17.4	87.0	171180	2	AC067986	AC067986 Homo sapi
35	17.4	87.0	176193	9	AC009707	AC009707 Homo sapi
36	17.4	87.0	182549	9	AF127019	AF127019 Homo sapi
c 37	17.4	87.0	190050	1	AL646059	AL646059 Ralstonia
38	17.4	87.0	310050	1	RME603642	AL603642 Rhizobium
c 39	17	85.0	35209	1	AF065159	AF065159 Bradyrhiz
40	16.8	84.0	404	9	HSA336371	AJ336371 Homo sapi
c 41	16.8	84.0	409	6	AX071181	AX071181 Sequence
42	16.8	84.0	429	9	HSFGFR2A01	AF097336 Homo sapi
43	16.8	84.0	609	1	AF330026	AF330026 Burkholde
44	16.8	84.0	625	9	HSA335299	AJ335299 Homo sapi
45	16.8	84.0	707	9	HSA336089	AJ336089 Homo sapi

ALIGNMENTS

RESULT 1
PAU38241/c
LOCUS PAU38241 3123 bp DNA linear BCT 05-OCT-1996
DEFINITION Pseudomonas aeruginosa orotate phophoribosyl transferase (pyrE), catabolite repression control protein (crc) and RNasePH (rph) genes, complete cds.
ACCESSION U38241 L12038
VERSION U38241.1 GI:1079660
KEYWORDS .
SOURCE Pseudomonas aeruginosa strain=PAO1.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 3123)
AUTHORS MacGregor,C.H., Arora,S.K., Hager,P.W., Dail,M.B. and Phibbs,P.V. Jr.
TITLE The nucleotide sequence of the Pseudomonas aeruginosa pyrE-crc-rph region and the purification of the crc gene product
JOURNAL J. Bacteriol. 178 (19), 5627-5635 (1996)
MEDLINE 96421989
REFERENCE 2 (bases 1 to 3123)
AUTHORS Hager,P.W. and Phibbs,P.V. Jr.
TITLE Direct Submission

```
  JOURNAL   Submitted (11-OCT-1995) Paul W. Hager, Microbiology & Immunology,
            East Carolina University, Greenville, NC 27858, USA
COMMENT     On Nov 29, 1995 this sequence version replaced gi:496210.
FEATURES             Location/Qualifiers
     source          1. .3123
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
                     /map="11 min"
     gene            complement(414. .1055)
                     /gene="pyrE"
     CDS             complement(414. .1055)
                     /gene="pyrE"
                     /EC_number="2.4.2.10"
                     /function="catalytic activity: orotidine-5'-phosphate +
                     pyrophosphate = orotate + 5-phospho-alpha-d-ribose
                     1-diphosphate; Method: conceptual translation supplied by
                     author."
                     /codon_start=1
                     /transl_table=11
                     /product="orotate phosphoribosyl transferase"
                     /protein_id="AAC44427.1"
                     /db_xref="GI:1079661"
                     /translation="MQAYQRDFIRFAIERGVLRFGEFTLKSGRTSPYFFNAGLFDSGL
                     ALARLGRFYAEAVIDSGIDFDVLFGPAYKGIPLAATTAVALAEQHQRDLPWCFNRKEA
                     KEHGEGGTLVGAPLSGRVLIIDDVITAGTAIREVMQIIDAQGARAAGVLIALNRQERG
                     KGELSAIQEVERDFGMPVVSIVSLEQVLEYLAEDAELKKHLPAVEAYRAQYGI"
     gene            1136. .1915
                     /gene="crc"
     CDS             1136. .1915
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAC44428.1"
                     /db_xref="GI:1079662"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(2394. .3113)
                     /gene="rph"
     CDS             complement(2394. .3113)
                     /gene="rph"
                     /note="similar to ribonuclease PH"
                     /codon_start=1
                     /transl_table=11
                     /product="RNase PH"
                     /protein_id="AAC44429.1"
                     /db_xref="GI:1079663"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
                     ESGVPRFLKGQGQGWLTAEYGMLRRSTGERNQREASRGKQGGRTLEIQRLIDRSLRAA
                     LDLSKLGENTLYIDCDVIQTDGGTRTASITGATVSLIDALAVLKKRVALKGNPLKQIV
                     AAVSVGIYQGVPVLDLDYLEDSAAESDLNVVMTDCRRFIEVQGTAEGAPFRPAELNAM
                     LELAQQGMQELVRTQRAALAE"
BASE COUNT      551 a    1020 c     989 g     563 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 1;  Length 3123;
  Best Local Similarity  100.0%;  Pred. No. 2.9e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcctg 20
          ||||||||||||||||||||
Db   1188 CCGCGCTCGGCCGCAGCCTG 1169


RESULT   2
AE004946/c
LOCUS       AE004946               13075 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 507 of 529 of the complete
            genome.
ACCESSION   AE004946 AE004091
VERSION     AE004946.1  GI:9951650
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .13075
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            71. .850
                     /gene="crc"
                     /note="PA5332"
     CDS             71. .850
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAG08717.1"
                     /db_xref="GI:9951651"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(909. .1280)
                     /gene="PA5333"
     CDS             complement(909. .1280)
                     /gene="PA5333"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08718.1"
                     /db_xref="GI:9951652"
                     /translation="MNDVSEEVAATPGQEARQWAMFCHFAAFIGLVFPFGSLLGPLIV
                     WQLKRESDPFVDAQGKEALNFQITVALAMFVSFLLMLVVIGFFLLGLVCLAALVLTII
                     AGIKANEGKAYRYPFCLRLIK"
     gene            complement(1327. .2046)
                     /gene="rph"
                     /note="PA5334"
     CDS             complement(1327. .2046)
                     /gene="rph"
                     /codon_start=1
                     /transl_table=11
                     /product="ribonuclease PH"
                     /protein_id="AAG08719.1"
                     /db_xref="GI:9951653"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
```

```
                ESGVPRFLKGQGQGWLTAEYGMLPRSTGERNQREASRGKQGGRTLEIQRLIGRSLRAA
                LDLSKLGENTLYIDCDVIQADGGTRTASITGATVALIDALAVLKKRGALKGNPLKQMV
                AAVSVGIYQGVPVLDLDYLEDSAAETDLNVVMTDAGGFIEVQGTAEGAPFRPAELNAM
                LELAQQGMQELFELQRAALAE"
gene            2227. .3090
                /gene="PA5335"
CDS             2227. .3090
                /gene="PA5335"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG08720.1"
                /db_xref="GI:9951654"
                /translation="MVHSMTAFARVEQAGTHGTLSWELRSVNHRYLEPHLRLPDAFRD
                LEGAVREALRQGLSRGKVECTLRFAEETAGKSLQVDQERARQLVAAAEGVAALIRQPA
                PLDPLAVLAWPGVLVADSADPQALNAAALEAFGQALEQLKAGRSREGLELAKLLNDRL
                DAMLVEVGNLRELVPTMLANQRQKILDRFAELKAELDPQRLEQELVLLAQKSDVAEEL
                DRLATHVGEVRRVLKAGGAAGRRLDFLMQELNREANTLGSKAFDPRSTQAAVNLKVLI
                EQMREQVQNIE"
gene            3149. .3760
                /gene="gmk"
                /note="PA5336"
CDS             3149. .3760
                /gene="gmk"
                /codon_start=1
                /transl_table=11
                /product="guanylate kinase"
                /protein_id="AAG08721.1"
                /db_xref="GI:9951655"
                /translation="MSGTLYIVSAPSGAGKTSLVKALLDAAPEVRVSVSHTTRGMRPG
                EVDGVNYHFTSREEFLAMLERNEFLEHAEVFGNLYGTSQRWVEKTLAEGLDLILEIDW
                QGAQQVRRLMPEAQSIFILPPSQEALRQRLTNRGQDSDEVIERRMREAVSEMSHYVEY
                DHLVINDDFAHALDDLKAIFRAROLRQDAQQQRHAELLGRLLA"
gene            3849. .4115
                /gene="rpoZ"
                /note="PA5337"
CDS             3849. .4115
                /gene="rpoZ"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase omega subunit"
                /protein_id="AAG08722.1"
                /db_xref="GI:9951656"
                /translation="MARVTVEDCLDNVDNRFELVMLATKRARQLATGGKEPKVAWEND
                KPTVVALREIASGLVDENVVQQEDIVEDEPLFAAFDDEANTEAL"
gene            4181. .6286
                /gene="spoT"
                /note="PA5338"
CDS             4181. .6286
                /gene="spoT"
                /codon_start=1
                /transl_table=11
                /product="guanosine-3',5'-bis(diphosphate)
                3'-pyrophosphohydrolase"
                /protein_id="AAG08723.1"
                /db_xref="GI:9951657"
                /translation="MPGIDAFADRLSSYLGPDQVNLVRRAYYYAEQAHDGQRRRSGEP
                YVTHPLAVSNILADMHMDHQSLMAAMLHDVIEDTGIAKEALTAQFGETVSELVDGVSK
                LTQMNFETKAEAQAENFQKMAMAMARDIRVILVKLADRLHNMRTLEVLSGEKRRRIAK
                ETLEIYAPIANRLGMHTMRVEFEDLGFKAMHPMRAERIRQAVKKARGNRREIVGKIQE
                SLVNCLAREGMEGQVIGREKHLYSIYQKMRGKRKAFNEIMDVYAFRIIVDKVDTCYRV
                LGAVHSLYKPFPGRFKDYIAIPKANGYQSLHTTLFGMHGVPIEIQIRTREMEEMANHG
                IAAHWLYKSNDETPKGTHARARQWVKGVLELQQRAGNSLEFIENVKIDLFPDEVYVFT
                PKGRIMELPKGSTAVDFAYAVHTDVGNSCIACRINRRLAPLSEPLQSGQTVEIVTAPG
                ARPNPAWLSFVVTGKARTHIRHALKLQRRSESINLGERLLNKTLTGFDSHLEKIPQER
                IQAVLAEYRMEVFEDLLEEIGLGNRMAYVVARRLLSSEGEEAPSAEGPLAIRGTEGLV
                LSYAKCCTPIPGDPIVGHLSAGKGMVVHLESCKNISEVRHNPEKCIQLSWAKDVTGEF
                NVELRVELEHQRGLIALLATSVNAADGNIEKISMDERDGRISVVQLVVSVHDRVHLAR
                VIKKLRALKGVIRITRLRS"
gene            6348. .6728
                /gene="PA5339"
CDS             6348. .6728
                /gene="PA5339"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG08724.1"
                /db_xref="GI:9951658"
                /translation="MTKTVIHTDNAPAAIGTYSQAIKAGNTVYVSGQIPLDPKTMELV
                EGFEEQTVQVFENLKAVIEASGGYLHDVAKLNIFLTDLSHFAKVNEIMGRYFTQPYPA
                RAAIGVAALPKGAQVEMDAIVVLE"
gene            6783. .7514
                /gene="PA5340"
CDS             6783. .7514
                /gene="PA5340"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG08725.1"
                /db_xref="GI:9951659"
                /translation="MRRVIFLAAAATLLAGCAGTADPSGTWINQAAIDAASKDGKLRE
                ALLAYGPNLEWKLDSKAGEATFSNGFELGEGTLSKSDDEHWKVAFYGDDNQESLELDG
                KELIQQASANGPEQRFRRLDPQPAASSPAGSGFERALYGSYLKGSWKIREGQGQGGKV
                EFQANGLVSGLPGAERYALCLAGDCAAMSGDNDSIWLQQGNRGRELLFSLDDDELQLF
                EAVNTAGANEMPSYVPGKRVWLLER"
gene            complement(7521. .8141)
                /gene="PA5341"
CDS             complement(7521. .8141)
                /gene="PA5341"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG08726.1"
                /db_xref="GI:9951660"
                /translation="MFAAFVLVASTHFAALLSPGPDFFLLVRAALLRGRRQADGVAAG
                IALANLLSMLLVLGLLASVPDSAHSALRWLQALGGLYFLWLAGQALLAQRRLEMPAQR
                DVPSRGGYLRGLRDGLLASSLNPKLPIFYAGLFGVLARFSLPGWALALCLAWMSLAVL
                CWDMALVRLLDRPRWRGWLQRRVGALDRLCGVLLLALGGWLVLAAL"
gene            complement(8208. .9008)
                /gene="PA5342"
CDS             complement(8208. .9008)
                /gene="PA5342"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG08727.1"
                /db_xref="GI:9951661"
                /translation="MAANRLLARPWLPGVELFHADFSGQAFGRHSHSAFAIGSILHGV
                GGYQCRGRRHALPAGTLSLMNPEEPHTGHAESERLVYRMLYVEETRLLGLLGRKQLPR
                GFGELNPADDGSVAAGLAWLAEDFQRSDALGLESRLLALLEQVFVRHGGLRPAAAAPR
                DGGTTARLRDYLEAHATEALGLDQLAQLLQRHPRHLIEAFRRAYGVPPHTYQLQCRVR
                QAKRLLLEERPLAELAQHLGFYDQAHFSTTFRRFTGVTPGQFRRSAKT"
gene            complement(9008. .9859)
                /gene="PA5343"
CDS             complement(9008. .9859)
```

Query Match 100.0%; Score 20; DB 1; Length 13075;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ccgcgctcggccgcagcctg 20
          ||||||||||||||||||||
Db    123 CCGCGCTCGGCCGCAGCCTG 104
```

Comp

```
RESULT  3
AC004396
LOCUS       AC004396               43903 bp    DNA     linear   HTG 13-MAR-2000
DEFINITION  Pseudomonas sp. chromosome Genomic clone cos3b, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC004396
VERSION     AC004396.12  GI:7230801
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Pseudomonas sp.
  ORGANISM  Pseudomonas sp.
            Bacteria; Proteobacteria.
```

```
REFERENCE   1  (bases 1 to 43903)
  AUTHORS   Roe,B.A.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43903)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:7229744.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    15232: contig of 15232 bp in length
            *   15233    15352: gap of unknown length
            *   15353    43903: contig of 28551 bp in length.
FEATURES             Location/Qualifiers
     source          1. .43903
                     /organism="Pseudomonas sp."
                     /db_xref="taxon:306"
                     /chromosome="Genomic"
                     /clone="cos3b"
BASE COUNT     8111 a  13520 c  13821 g   8330 t    121 others
ORIGIN

  Query Match           92.0%;  Score 18.4;  DB 2;  Length 43903;
  Best Local Similarity 95.0%;  Pred. No. 7.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcctg 20
          || ||||||||||||||||||
Db  37847 CCACGCTCGGCCGCAGCCTG 37866


RESULT   4
HSA335678/c
LOCUS       HSA335678               650 bp    DNA     linear   PRI 01-OCT-2001
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NR1-DK7C.
ACCESSION   AJ335678
VERSION     AJ335678.1  GI:15880096
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
            Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
            Muravenko,O.V., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L.,
            Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     Analysis of NotI flanking sequences: a new tool for gene discovery
            and verification of the human genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 650)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NR1-DK7C"
BASE COUNT      114 a    150 c    191 g    162 t     33 others
ORIGIN

  Query Match           87.0%;  Score 17.4;  DB 9;  Length 650;
  Best Local Similarity 94.7%;  Pred. No. 4.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||| ||||||||
Db    110 CGCGCTCGGCGGCAGCCTG 92


RESULT   5
HSA337098/c
LOCUS       HSA337098               690 bp    DNA     linear   PRI 01-OCT-2001
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NR5-CC20R.
ACCESSION   AJ337098
VERSION     AJ337098.1  GI:15881516
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
            Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
            Muravenko,O.V., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L.,
            Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     Analysis of NotI flanking sequences: a new tool for gene discovery
            and verification of the human genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 690)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NR5-CC20R"
BASE COUNT      148 a    214 c    200 g    128 t
ORIGIN

  Query Match           87.0%;  Score 17.4;  DB 9;  Length 690;
  Best Local Similarity 94.7%;  Pred. No. 4.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          ||| ||||||||||||||||
Db     35 CGCACTCGGCCGCAGCCTG 17


RESULT   6
HS2NRG12
LOCUS       HS2NRG12                970 bp    DNA     linear   PRI 01-FEB-2000
DEFINITION  Homo sapiens alternatively spliced neuregulin 2 (NRG2) gene, exon
            12 and complete cds.
ACCESSION   AF119162
VERSION     AF119162.1  GI:6840971
KEYWORDS    .
SEGMENT     12 of 12
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 970)
  AUTHORS   Ring,H.Z., Chang,H., Guilbot,A., Brice,A., LeGuern,E. and
            Francke,U.
  TITLE     The human neuregulin 2 (NRG2) gene: cloning, mapping and evaluation
```

```
          as a candidate for the autosomal recessive form of
          Charcot-Marie-Tooth disease linked to 5q
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 970)
  AUTHORS Ring,H.Z., Chang,H. and Francke,U.
  TITLE   Direct Submission
  JOURNAL Submitted (12-JAN-1999) Department of Genetics and HHMI, Stanford
          University School of Medicine, B203 Beckman Center, Stanford, CA
          94305, USA
FEATURES             Location/Qualifiers
     source          1. .970
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q23-q33"
     mRNA            join(AF119151.1:<313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119156.1:123. .181,AF119157.1:30. .53,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .>851)
                     /gene="NRG2"
                     /note="alternatively spliced"
                     /product="neuregulin 2 isoform 4"
     mRNA            join(AF119151.1:<313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119156.1:123. .181,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .>851)
                     /gene="NRG2"
                     /note="alternatively spliced"
                     /product="neuregulin 2 isoform 2"
     mRNA            join(AF119151.1:<313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119155.1:111. .187,AF119157.1:30. .53,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .>851)
                     /gene="NRG2"
                     /note="alternatively spliced"
                     /product="neuregulin 2 isoform 3"
     mRNA            join(AF119151.1:<313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119155.1:111. .187,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .>851)
                     /gene="NRG2"
                     /note="alternatively spliced"
                     /product="neuregulin 2 isoform 1"
     gene            order(AF119151.1:<313. .1054,AF119152.1:1. .419,
                     AF119153.1:1. .493,AF119154.1:1. .350,AF119155.1:1. .326,
                     AF119156.1:1. .310,AF119157.1:1. .258,AF119158.1:1. .259,
                     AF119159.1:1. .300,AF119160.1:1. .209,AF119161.1:1. .522,
                     1. .>851)
                     /gene="NRG2"
     CDS             join(AF119151.1:313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119155.1:111. .187,AF119157.1:30. .53,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .851)
                     /gene="NRG2"
                     /note="alternatively spliced; a member of the neuregulin
                     family of growth and differentiation factors"
                     /codon_start=1
                     /product="neuregulin 2 isoform 3"
                     /protein_id="AAF28850.1"
                     /db_xref="GI:6840975"
                     /translation="MRQVCCSALPPPPLEKGRCSSYSDSSSSSSERSSSSSSSSSESG
                     SSSRSSSNNSSISRPAAPPEPRPQQQPQPRSPAARRAAARSRAAAAGGMRRDPAPGFS
                     MLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVAL
                     VKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFAPLDTNG
                     KNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKE
                     LNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVSTT
                     LSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP
                     DPKQKHLGFELKEAEELYQKRVLTITGICVALLVVGIVCVVAYCKTKKQRKQMHNHLR
                     QNMCPAHQNRSLANGPSHPRLDPEEIQMADYISKNVPATDHVIRRETETTFSGSHSCS
                     PSHHCSTATPTSSHRHESHTWSLERSESLTSDSQSGIMLSSVGTSKCNSPACVEARAR
                     RAAAYNLEERRRATAPPYHDSVDSLRDSPHSERYVSALTTPARLSPVDFHYSLATQVP
                     TFEITSPNSAHAVSLPPAAPISYRLAEQQPLLRHPAPPGPGPGPGPGPGPGADMQRSY
                     DSYYYPAAGPGPRRGTCALGGSLGSLPASPFRIPEDDEYETTQECAPPPPPRPRARGA
                     SRRTSAGPRRWRRSRLNGLAAQRARAARDSLSLSSGSGGGSASASDDDADDADGALAA
                     ESTPFLGLRGAHDALRSDSPPLCPAADSRTYYSLDSHSTRASSRHSRGPPPRAKQDSA
                     PL"
     CDS             join(AF119151.1:313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119156.1:123. .181,AF119157.1:30. .53,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .851)
                     /gene="NRG2"
                     /note="alternatively spliced; a member of the neuregulin
                     family of growth and differentiation factors"
                     /codon_start=1
                     /product="neuregulin 2 isoform 4"
                     /protein_id="AAF28851.1"
                     /db_xref="GI:6840976"
                     /translation="MRQVCCSALPPPPLEKGRCSSYSDSSSSSSERSSSSSSSSSESG
                     SSSRSSSNNSSISRPAAPPEPRPQQQPQPRSPAARRAAARSRAAAAGGMRRDPAPGFS
                     MLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVAL
                     VKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFAPLDTNG
                     KNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKE
                     LNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVSTT
                     LSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP
                     DPKQKHLGFELKEAEELYQKRVLTITGICVALLVVGIVCVVAYCKTKKQRKQMHNHLR
                     QNMCPAHQNRSLANGPSHPRLDPEEIQMADYISKNVPATDHVIRRETETTFSGSHSCS
                     PSHHCSTATPTSSHRHESHTWSLERSESLTSDSQSGIMLSSVGTSKCNSPACVEARARRAAAYN
                     LEERRRATAPPYHDSVDSLRDSPHSERYVSALTTPARLSPVDFHYSLATQVPTFEITS
                     PNSAHAVSLPPAAPISYRLAEQQPLLRHPAPPGPGPGPGPGPGADMQRSYDSYYYP
                     AAGPGPRRGTCALGGSLGSLPASPFRIPEDDEYETTQECAPPPPPRPRARGASRRTSA
                     GPRRWRRSRLNGLAAQRARAARDSLSLSSGSGGGSASASDDDADDADGALAAESTPFL
                     GLRGAHDALRSDSPPLCPAADSRTYYSLDSHSTRASSRHSRGPPPRAKQDSAPL"
     CDS             join(AF119151.1:313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119156.1:123. .181,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .851)
                     /gene="NRG2"
                     /note="alternatively spliced; a member of the neuregulin
                     family of growth and differentiation factors"
                     /codon_start=1
                     /product="neuregulin 2 isoform 2"
                     /protein_id="AAF28849.1"
                     /db_xref="GI:6840974"
                     /translation="MRQVCCSALPPPPLEKGRCSSYSDSSSSSSERSSSSSSSSSESG
                     SSSRSSSNNSSISRPAAPPEPRPQQQPQPRSPAARRAAARSRAAAAGGMRRDPAPGFS
                     MLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVAL
                     VKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFAPLDTNG
                     KNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKE
                     LNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVSTT
                     LSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCQQFAMVNFSKA
                     EELYQKRVLTITGICVALLVVGIVCVVAYCKTKKQRKQMHNHLRQNMCPAHQNRSLAN
                     GPSHPRLDPEEIQMADYISKNVPATDHVIRRETETTFSGSHSCSPSHHCSTATPTSSH
                     RHESHTWSLERSESLTSDSQSGIMLSSVGTSKCNSPACVEARARRAAAYNLEERRRAT
                     APPYHDSVDSLRDSPHSERYVSALTTPARLSPVDFHYSLATQVPTFEITSPNSAHAVS
                     LPPAAPISYRLAEQQPLLRHPAPPGPGPGPGPGPGPGADMQRSYDSYYYPAAGPGPRR
                     GTCALGGSLGSLPASPFRIPEDDEYETTQECAPPPPPRPRARGASRRTSAGPRRWRRS
                     RLNGLAAQRARAARDSLSLSSGSGGGSASASDDDADDADGALAAESTPFLGLRGAHDA
                     LRSDSPPLCPAADSRTYYSLDSHSTRASSRHSRGPPPRAKQDSAPL"
     CDS             join(AF119151.1:313. .1012,AF119152.1:51. .222,
                     AF119153.1:230. .348,AF119154.1:100. .220,
                     AF119155.1:111. .187,AF119158.1:132. .234,
                     AF119159.1:82. .205,AF119160.1:46. .173,AF119161.1:111. .338,
                     71. .851)
                     /gene="NRG2"
                     /note="alternatively spliced; a member of the neuregulin
                     family of growth and differentiation factors"
                     /codon_start=1
                     /product="neuregulin 2 isoform 1"
                     /protein_id="AAF28848.1"
```

```
                     /db_xref="GI:6840973"
                     /translation="MRQVCCSALPPPPLEKGRCSSYSDSSSSSSERSSSSSSSSESG
                     SSSRSSSNNSSISRPAAPPEPRPQQQPQPRSPAARRAAARSRAAAAGGMRRDPAPGFS
                     MLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVAL
                     VKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFAPLDTNG
                     KNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKE
                     LNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVSTT
                     LSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP
                     DPKQKAEELYQKRVLTITGICVALLVVGIVCVVAYCKTKKQRKQMHNHLRQNMCPAHQ
                     NRSLANGPSHPRLDPEEIQMADYISKNVPATDHVIRRETETTFSGSHSCSPSHHCSTA
                     TPTSSHRHESHTWSLERSESLTSDSQSGIMLSSVGTSKCNSPACVEARARRAAAYNLE
                     ERRRATAPPYHDSVDSLRDSPHSERYVSALTTPARLSPVDFHYSLATQVPTFEITSPN
                     SAHAVSLPPAAPISYRLAEQQPLLRHPAPPGPGPGPGPGPGPGPGADMQRSYDSYYYPAA
                     GPGPRRGTCALGGSLGSLPASPFRIPEDDEYETTQECAPPPPPRPRARGASRRTSAGP
                     RRWRRSRLNGLAAQRARAARDSLSLSSGSGGGSASASDDDADDADGALAAESTPFLGL
                     RGAHDALRSDSPPLCPAADSRTYYSLDSHSTRASSRHSRGPPPRAKQDSAPL"
     exon            71. .>851
                     /gene="NRG2"
                     /number=12
BASE COUNT      145 a    396 c    319 g    110 t
ORIGIN


  Query Match             87.0%;  Score 17.4;  DB 9;  Length 970;
  Best Local Similarity   94.7%;  Pred. No. 3.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          ||||||||||| ||||||||
Db    371 CGCGCTCGGCGGCAGCCTG 389


RESULT   7
AB063178
LOCUS       AB063178               1889 bp    mRNA    linear   PLN 28-NOV-2001
DEFINITION  Hordeum vulgare BBD2 mRNA for betaine aldehyde dehydrogenase,
            complete cds.
ACCESSION   AB063178
VERSION     AB063178.1  GI:15147870
KEYWORDS    .
SOURCE      Hordeum vulgare (cultivar:Haruna-nijyo) cDNA to mRNA, clone:BBD2.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (sites)
  AUTHORS   Nakamura,T., Nomura,M., Mori,H., Jagendorf,A.T., Ueda,A. and
            Takabe,T.
  TITLE     An isozyme of betaine aldehyde dehydrogenase in barley
  JOURNAL   Plant Cell Physiol. 42 (10), 1088-1092 (2001)
  MEDLINE   21530294
REFERENCE   2  (bases 1 to 1889)
  AUTHORS   Nakamura,T. and Takabe,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Tetsuko Takabe, Graduate School of
            Bioagricultural Sciences, Nagoya University; Furo-cho, Chikusa,
            Nagoya, Aichi 464-8601, Japan
            (E-mail:h44854a@nucc.cc.nagoya-u.ac.jp, Tel:81-52-789-5209,
            Fax:81-52-789-5209)
FEATURES             Location/Qualifiers
     source          1. .1889
                     /organism="Hordeum vulgare"
                     /cultivar="Haruna-nijyo"
                     /db_xref="taxon:4513"
                     /clone="BBD2"
     gene            47. .1558
                     /gene="BBD2"
     CDS             47. .1558
                     /gene="BBD2"
                     /note="dicotyledon type"
                     /codon_start=1
                     /product="betaine aldehyde dehydrogenase"
                     /protein_id="BAB62846.1"
                     /db_xref="GI:15147871"
                     /translation="MVAPAKIPQRQLFIDGEWRAPALGRRLPVINPTTEVSIGEIPAG
                     TSEDVDAAVAAARAALKRNRGRDWSRAPGAVRAKYLRAIAAKMIERKSDLARLEALDC
                     GKPLDEAAWDMDDVAGCFEFFAGHAEALDKRQNAAVALPENFKCHLKKEPIGVVALIT
                     PWNYPLLMAVWKVAPALAAGCTAVLKPSELASVTCLELGDVCKEVGLPSGVLNIVTGL
                     GNEAGAPLSSHPDVDKVAFTGSYATGQKIMVAAAPTVKPVTLELGGKSPIVVFDDVDI
                     DKAVEWTLFGCFWTNGQICSATSRLLIHKNIAKEFVDRMVAWSKNIKVSDPLEEGCRL
                     GPVVSEGQYEKIKKFVANAKTEGATILTGGVRPKHLEKGFFIEPTIITDINTSMEIWR
                     EEVFGPVLCVKEFSTEEEAIELANDTHYGLAGAVISGDRERCQRLAEEIEAGCIWVNC
                     SQPCFCQAPWGGNKRSGFGRELGEGGIDNYLSIKQVTEYTSDAPWGWYKAPAN"
BASE COUNT      455 a    443 c    534 g    457 t
ORIGIN


  Query Match             87.0%;  Score 17.4;  DB 8;  Length 1889;
  Best Local Similarity   94.7%;  Pred. No. 3.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          |||||||||||||| ||||
Db    108 CCGCGCTCGGCCGCCGCCT 126


RESULT   8
AR098155
LOCUS       AR098155               2268 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 31 from patent US 6074841.
ACCESSION   AR098155
VERSION     AR098155.1  GI:12807412
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2268)
  AUTHORS   Gearing,D.P. and Busfield,S.J.
  TITLE     Don-1 gene and polypeptides and uses therefor
  JOURNAL   Patent: US 6074841-A 31 13-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .2268
                     /organism="unknown"
BASE COUNT      502 a    734 c    701 g    331 t
ORIGIN


  Query Match             87.0%;  Score 17.4;  DB 6;  Length 2268;
  Best Local Similarity   94.7%;  Pred. No. 3.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          ||||||||||| ||||||||
Db   1532 CGCGCTCGGCGGCAGCCTG 1550


RESULT   9
AR116627
LOCUS       AR116627               2268 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 31 from patent US 6133423.
ACCESSION   AR116627
VERSION     AR116627.1  GI:14096949
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2268)
  AUTHORS   Gearing,D.P. and Busfield,S.J.
  TITLE     Don-1 gene and polypeptides and uses therefor
  JOURNAL   Patent: US 6133423-A 31 17-OCT-2000;
FEATURES             Location/Qualifiers
     source          1. .2268
                     /organism="unknown"
BASE COUNT      502 a    734 c    701 g    331 t
ORIGIN
```

```
  Query Match             87.0%;  Score 17.4;  DB 6;  Length 2268;
  Best Local Similarity   94.7%;  Pred. No. 3.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||| ||||||||
Db   1532 CGCGCTCGGCGGCAGCCTG 1550


RESULT  10
AB005060
LOCUS       AB005060               3020 bp    mRNA    linear   PRI 20-NOV-1997
DEFINITION  Homo sapiens mRNA for NTAK, complete cds.
ACCESSION   AB005060
VERSION     AB005060.1  GI:2626738
KEYWORDS    NTAK.
SOURCE      Homo sapiens neuroblastoma cell_line:SK-N-SH cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3020)
  AUTHORS   Ishiguro,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1997) Hiroshi Ishiguro, Fujita Health University,
            ICMS; 1-98, kutsukake-cho, Toyoake, Aichi 470-11, Japan
            (E-mail:hishi@fujita-hu.ac.jp, Tel:0562-93-9393, Fax:0562-93-8831)
REFERENCE   2  (sites)
  AUTHORS   Higashiyama,S., Horikawa,M., Yamada,K., Ichino,N., Nakano,N.,
            Nakagawa,T., Miyagawa,J., Matsushita,N., Nagatsu,T., Taniguchi,N.
            and Ishiguro,H.
  TITLE     A novel brain-derived member of the epidermal growth factor family
            that interacts with ErbB3 and ErbB4
  JOURNAL   J. Biochem. 122 (3), 675-680 (1997)
  MEDLINE   98006324
FEATURES             Location/Qualifiers
     source          1. .3020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="SK-N-SH"
                     /cell_type="neuroblastoma"
     CDS             226. .2778
                     /codon_start=1
                     /product="NTAK"
                     /protein_id="BAA23417.1"
                     /db_xref="GI:2626739"
                     /translation="MRQVCCSALPPPPLEKGRCSSYSDSSSSSSERSSSSSSSSSESG
                     SSSRSSSNNSSISRPAAPPEPRPQQQPQPRSPAARRAAARSRAAAAGGMRRDPAPGFS
                     MLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVAL
                     VKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFAPLDTNG
                     KNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKE
                     LNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVRGRLYVNSVSTT
                     LSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMP
                     DPKQKAEELYQKRVLTITGICVALLVVGIVCVVAYCKTKKQRKQMHNHLRQNMCPAHQ
                     NRSLANGPSHPRLDPEEIQMADYISKNVPATDHVIRRETETTFSGSHSCSPSHHCSTA
                     TPTSSHRHESHTWSLERSESLTSDSQSGIMLSSVGTSKCNSPACVEARARRAAAYNLE
                     ERRRATAPPYHDSVDSLRDSPHSERYVSALTTPARLSPVDFHYSLATQVPTFEITSPN
                     SAHAVSLPPAAPISYRLAEQQPLLRHPAPPGPGPGPGPGPGPGADMQRSYDSYYYPAA
                     GPGPRRGTCALGGSLGSLPASPFRIPEDDEYETTQECAPPPPPRPRARGASRRTSAGP
                     RRWRRSRLNGLAAQRARAARDSLSLSSGSGGGSASASDDDADDADGALAAESTPFLGL
                     RGAHDALRSDSPPLCPAADSRTYYSLDSHSTRASSRHSRGPPPRAKQDSAPL"
     polyA_site      3020
                     /note="39 A nucleotides"
BASE COUNT      615 a   1015 c    937 g    453 t
ORIGIN


  Query Match             87.0%;  Score 17.4;  DB 9;  Length 3020;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||| ||||||||

Db   2298 CGCGCTCGGCGGCAGCCTG 2316


RESULT  11
DVU16723/c
LOCUS       DVU16723               3119 bp    DNA     linear   BCT 20-JUL-1995
DEFINITION  Desulfovibrio vulgaris dissimilatory sulfite reductase alpha
            (dsvA), dissimilatory sulfite reductase beta (dsvB) and dsvD genes,
            complete cds.
ACCESSION   U16723
VERSION     U16723.1  GI:902746
KEYWORDS    .
SOURCE      Desulfovibrio vulgaris.
  ORGANISM  Desulfovibrio vulgaris
            Bacteria; Proteobacteria; delta subdivision; Desulfovibrio.
REFERENCE   1  (bases 1 to 3119)
  AUTHORS   Karkhoff-Schweizer,R.R., Huber,D.P. and Voordouw,G.
  TITLE     Conservation of the genes for dissimilatory sulfite reductase from
            Desulfovibrio vulgaris and Archaeoglobus fulgidus allows their
            detection by PCR
  JOURNAL   Appl. Environ. Microbiol. 61 (1), 290-296 (1995)
  MEDLINE   95193998
REFERENCE   2  (bases 1 to 3119)
  AUTHORS   Voordouw,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1994) Gerrit Voordouw, Department of Biological
            Sciences, the University of Calgary, 2500 University Drive NW,
            Calgary, Alberta T2N 1N4, Canada
COMMENT     Dissimilatory sulfite reductase catalyzes the reduction of sulfite
            to sulfide.  This is the terminal oxidation reaction in sulfate
            respiration, a process catalyzed by the sulfate-reducing bacteria.
FEATURES             Location/Qualifiers
     source          1. .3119
                     /organism="Desulfovibrio vulgaris"
                     /strain="Hildenborough"
                     /sub_species="vulgaris"
                     /db_xref="taxon:881"
                     /clone="8303"
                     /clone_lib="NCIMB"
     -35_signal      42. .47
                     /note="putative"
     -10_signal      66. .71
                     /note="putative"
     RBS             223. .227
                     /gene="dsvA"
                     /note="putative"
     gene            223. .1548
                     /gene="dsvA"
     CDS             235. .1548
                     /gene="dsvA"
                     /function="catalyst for the reduction of sulfite to
                     sulfide"
                     /note="similar to archaebacterium Archaeoglobus fulgidus
                     sulfite reductase alpha, PIR Accession Number S27478"
                     /codon_start=1
                     /transl_table=11
                     /product="dissimilatory sulfite reductase alpha"
                     /protein_id="AAA70107.1"
                     /db_xref="GI:902747"
                     /translation="MAKHATPKLDQLESGPWPSFVSDIKQEAAYRAANPKGLDYQVPV
                     DCPEDLLGVLELSYDEGETHWKHGGIVGVFGYGGGVIGRYCDQPEKFPGVAHFHTVRV
                     AQPSGKYYSADYLRQLCDIWDLRGSGLTNMHGSTGDIVLLGTQTPQLEEIFFELTHNL
                     NTDLGGSGSNLRTPESCLGKSRCEFACYDSQAACYELTMEYQDELHRPAFPYKFKFKF
                     DACPNGCVASIARSDFSVIGTWKDDIKIDAEAVKAYVAGEFKPNAGAHSGRDWGKFDI
                     EAEVVNRCPSKCMKWDGSKLSIDNKECVRCMHCINTMPRALHIGDERGASILCGAKAP
                     ILDGAQMGSLLVPFVAAEEPFDEIKEVVEKIWDWWMEEGKNRERLGETMKRLSFQKLL
                     EVTEIAPVPQHVKEPRTNPYIFFKEEEVPGGWDRDITEYRKRHLR"
     RBS             1550. .1555
                     /gene="dsvB"
                     /note="putative"
     gene            1550. .2712
                     /gene="dsvB"
     CDS             1567. .2712
```

```
                     /gene="dsvB"
                     /function="catalyst for the reduction of sulfite to
                     sulfide"
                     /note="similar to archaebacterium Archaeoglobus fulgidus
                     sulfite reductase beta, PIR Accession Number S27379"
                     /codon_start=1
                     /transl_table=11
                     /product="dissimilatory sulfite reductase beta"
                     /protein_id="AAA70108.1"
                     /db_xref="GI:902748"
                     /translation="MAFISSGYNPEKPMANRITDIGPRKFDEFFPPVIAKNFGSWLYH
                     EILEPGVLMHVAESGDKVYTVRVGAARLMSITHIREMCDIADKYCGGHLRFTTRNNVE
                     FMVADEASLKALKEDLASRKFDGGSLKFPIGGTGAGVSNIVHTQGWVHCHTPATDASG
                     PVKAIMDEVFEDFQSMRLPAPVRISLACCINMCGAVHCSDIGVVGIHRKPPMIDHEWT
                     DQLCEIPLAVASCPTAAVRPTKLEIGDKKVNTIAIKNERCMYCGNCYTMCPALPISDG
                     EGDGVVIMVGGKVSNRISMPKFSKVVVAYIPNEPPRWPSLTKTIKHIIEVYSANAYKY
                     ERLGEWAERIGWERFFSLTGLEFSHHLIDDFRDPAYYTWRQSTQFKF"
     gene            2758. .3007
                     /gene="dsvD"
     RBS             2758. .2762
                     /gene="dsvD"
                     /note="putative"
     CDS             2771. .3007
                     /gene="dsvD"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA70109.1"
                     /db_xref="GI:902749"
                     /translation="MEEAKQKVVDFLNSKSGSKSKFYFNDFTDLFPDMKQREVKKILT
                     ALVNDEVLEYWSSGSTTMYGLKGAGKQAAAEHED"
BASE COUNT      670 a   1012 c    872 g    565 t
ORIGIN


  Query Match           87.0%;  Score 17.4;  DB 1;  Length 3119;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 cgcgctcggccgcagcctg 20
           || ||||||||||||||||
Db    2999 CGTGCTCGGCCGCAGCCTG 2981


RESULT  12
OSA427974
LOCUS       OSA427974               3777 bp    DNA     linear   PLN 28-JAN-2002
DEFINITION  Oryza sativa HAK3 gene for putative potasium transporter, exons
            1-6.
ACCESSION   AJ427974
VERSION     AJ427974.1  GI:18250693
KEYWORDS    HAK3 gene; potasium transporter.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Banuelos,M.A., Garciadeblas,B. and Rodriguez-Navarro,A.
  TITLE     Inventory of HAK transporters in rice
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3777)
  AUTHORS   Rodriguez-Navarro,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2002) Rodriguez-Navarro A., Biotecnologia,
            ETSIAgronomos UPM, Ciudad Universitaria s/n, Madrid, E-28040, SPAIN
COMMENT     This sequence was downloaded from Monsanto's rice genome database.
            Monsanto has granted permission to deposit this sequence in EMBL as
            part of the publication process.
FEATURES             Location/Qualifiers
     source          1. .3777
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
     gene            147. .3281
                     /gene="HAK3"
     exon            <147. .158
                     /gene="HAK3"
                     /number=1
     CDS             join(147. .158,504. .1002,1102. .1413,1497. .1549,1692. .1809,
                     1912. .3281)
                     /gene="HAK3"
                     /codon_start=1
                     /product="putative potasium transporter"
                     /protein_id="CAD20995.1"
                     /db_xref="GI:18250694"
                     /translation="MSEGGHWRSYYRHVLLLAYQSCGVVYGDLSTSPLYVYKSTFIIG
                     SLRRFQDEEIVFGVFSLVFWTLTLIPLLKYVFIVLAADDNGEGGTFALYSLLVRHAKF
                     SLMPNQEAADEELTSYYRPGYAPQETPILTALRRFLENHRKSRTFLLVTVLFGASLVI
                     GDGVLTPPMSVLSSFSGLQVHSTALTSGEVEILSCTVLVCLFMVQHWGTHRVAFLFAP
                     VVIVWLLLLGALGVYNIVVWNPRVLRALSPYYLVRFFQHTGKDGWISLGGILLSMTGT
                     EAMYADLGHFTAASIRVAFVGLIYPCLVLQYMGQAAFLSKSPHCDIHFVFFESIPTGI
                     FWPVLVIATLAAIVGSQAVISATFSIVRQCTALGCFPRVKIVHTSRRIHGQIYSPEIN
                     WILMLLCIAVTMGLRDTTLIGNAYGMACAGVMLVTTLLMALVIVFVWQYSCLVAALFL
                     VAFGVVEAVYLSAALMKVPQGGWLPLVLSLVFVAVMYVWHYGTRRKHQFDVQNKVSLR
                     WIHALGPSLGIVRVPGIGIIYSELATGVPAIFSHFVTNLPAFHQVLVFICVKAVPVPH
                     VRDEERHLVGRIGPREFRMYRCVVRHGYKDVLAEDTDFENDLVLRIAEFVQMEADFDQ
                     RCSISDDGVVASVEVEGRMAVVPRPSDLARTGLLMREPGEEESVVARAAAAAKPESLI
                     HSMHTMHEAESPGFASRRRVRFEVANQHTDPRVKEELSALVEAKHAGVAYIMGHSYIK
                     ARKSSSVFKKFAVNVAYAFLRKNCRGPGLVLNIPHISLIEVGMIYYV"
     intron          159. .503
                     /gene="HAK3"
                     /number=1
     exon            504. .1002
                     /gene="HAK3"
                     /number=2
     intron          1003. .1101
                     /gene="HAK3"
                     /number=2
     exon            1102. .1413
                     /gene="HAK3"
                     /number=3
     intron          1414. .1496
                     /gene="HAK3"
                     /number=3
     exon            1497. .1549
                     /gene="HAK3"
                     /number=4
     intron          1550. .1691
                     /gene="HAK3"
                     /number=4
     exon            1692. .1809
                     /gene="HAK3"
                     /number=5
     intron          1810. .1911
                     /gene="HAK3"
                     /number=5
     exon            1912. .>3281
                     /gene="HAK3"
                     /number=6
BASE COUNT      794 a   1045 c   1052 g    886 t
ORIGIN


  Query Match           87.0%;  Score 17.4;  DB 8;  Length 3777;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 cgcgctcggccgcagcctg 20
           |||||||||| |||||||
Db    2450 CGCGCTCGGCCCCAGCCTG 2468


RESULT  13
A06936/c
LOCUS       A06936                  4160 bp    DNA     linear   PAT 04-OCT-1993
```

```
DEFINITION  H.sapiens fur gene and c-fes/fps gene, exon Z and exons 1, 2.
ACCESSION   A06936
VERSION     A06936.1  GI:490064
KEYWORDS    fur gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4160)
  AUTHORS   Van de Ven,W.J.M., Roebroek,A.J.M. and Schalken,J.A.
  TITLE     Recombinant DNA and cDNA, mRNA, protein, antibodies, and a method
            of detecting tumor cells
  JOURNAL   Patent: EP 0246709-A 3 25-NOV-1987;
            Stichting Katholieke Universiteit
FEATURES             Location/Qualifiers
     source          1. .4160
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            281. .2271
                     /gene="fur"
     exon            281. .2271
                     /gene="fur"
                     /note="exon z"
     intron          2272. .3344
                     /number=1
     gene            3345. .4071
                     /gene="c-fes/fps"
     exon            3345. .3357
                     /gene="c-fes/fps"
                     /number=1
     intron          3358. .3849
                     /gene="c-fes/fps"
                     /number=2
     exon            3850. .4071
                     /gene="c-fes/fps"
                     /number=2
BASE COUNT      668 a   1323 c   1324 g    844 t      1 others
ORIGIN
```

```
  Query Match             87.0%;  Score 17.4;  DB 6;  Length 4160;
  Best Local Similarity   94.7%;  Pred. No. 2.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||
Db   3169 CCGCGCTGGGCCGCAGCCT 3151
```

```
RESULT  14
E01406/c
LOCUS       E01406                  4160 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence of 4.16kb human Bam H1/Bg 1 II.
ACCESSION   E01406
VERSION     E01406.1  GI:2169662
KEYWORDS    JP 1987285793-A/1.
SOURCE      Human herpesvirus 3.
  ORGANISM  Human herpesvirus 3
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 4160)
  AUTHORS   Buiremu,J.M.F.F., Antoniusu,Y.M.R.R. and Yakobusu,A.S. .
  TITLE     METHOD FOR DETECTING RECOMBINANT DNA, CDNA, MRNA, PROTEIN, ANTIBODY
            AND TUMOR CELL
  JOURNAL   Patent: JP 1987285793-A 1 11-DEC-1987;
            SUTEIKITEINGU KASORIIKE UNIV
COMMENT     PN   JP 1987285793-A/1
            PD   11-DEC-1987
            PF   20-MAY-1987 JP 1987123633
            PR   20-MAY-1986 NL 86   8601271
            PI   BUIREMU JIYAN MARIE FUAN DOU FUEN,
            PI   ANTONIUSU YOHANESU MARIA ROEBUROEKU,
            PI   YAKOBUSU ANTONIUSU SUKARUKEN
            PC   C12N15/00,A61K39/395,C07H21/04,C07K13/00,C07K15/04,C12Q1/00,
            PC   C12Q1/68,
            PC   G01N33/574//C12P21/00,G01N33/577;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: cell_line=KG-1;
            FH   Key             Location/Qualifiers
            FH
            FT   intron          1. .280
            FT   exon            281. .2271
            FT   intron          2272. .3345
            FT   exon            3346. .3357
            FT   intron          3358. .3849
            FT   exon            3850. .4071
            FT   intron          4072. .4160
            FT   CDS             281. .2271
            FT                   /note='fur gene'
            FT   CDS             join(3346. .3357,3850. .4071)
            FT                   /product='fes/fps protein'.
FEATURES             Location/Qualifiers
     source          1. .4160
                     /organism="Human herpesvirus 3"
                     /db_xref="taxon:10335"
BASE COUNT      669 a   1323 c   1324 g    844 t
ORIGIN
```

```
  Query Match             87.0%;  Score 17.4;  DB 6;  Length 4160;
  Best Local Similarity   94.7%;  Pred. No. 2.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||
Db   3169 CCGCGCTGGGCCGCAGCCT 3151
```

```
RESULT  15
AX097467/c
LOCUS       AX097467                7335 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 32 from Patent WO0118248.
ACCESSION   AX097467
VERSION     AX097467.1  GI:13514055
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 7335)
  AUTHORS   Whiteley,M., Lee,K.M., Greenberg,E.P. and Muh,U.
  TITLE     Quorum sensing signaling in bacteria
  JOURNAL   Patent: WO 0118248-A 32 15-MAR-2001;
            University of Iowa Research Foundation Inc. (US) ; Quorum Sciences
            Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .7335
                     /organism="Pseudomonas aeruginosa"
                     /db_xref="taxon:287"
BASE COUNT     1184 a   2256 c   2625 g   1270 t
ORIGIN
```

```
  Query Match             87.0%;  Score 17.4;  DB 6;  Length 7335;
  Best Local Similarity   94.7%;  Pred. No. 2.5e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||||||||||| |
Db   4450 CGCGCTCGGCCGCAGCCAG 4432
```

Search completed: June 23, 2002, 06:31:23
Job time: 63815 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 22, 2002, 16:27:48 ; Search time 652.13 Seconds
(without alignments)
52.656 Million cell updates/sec

Title: US-09-747-514A-1
Perfect score: 20
Sequence: 1 ccgcgctcggccgcagcctg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		17.4	87.0	2268	19	AAV17816	Homo sapiens don-1
c 2		17.4	87.0	7335	22	AAF81367	Quorum sensing con
3		17.4	87.0	28804	17	AAT37329	Sphingan biosynthe
4		17.4	87.0	28804	18	AAT92474	Sphingomonas genus
5		17.4	87.0	28804	20	AAV99812	Sphingomonas S88 s
6		17.4	87.0	28804	20	AAV81474	Chromosomal fragme
c 7		16.8	84.0	409	22	AAF65897	Novel human polynu
c 8		16.8	84.0	734	22	AAH05140	Human cDNA clone (
9		16.8	84.0	819	19	AAV12277	Maleate cis-trans
c 10		16.8	84.0	1877	22	AAH46831	Human immunoglobul
c 11		16.8	84.0	2934	23	AAS84820	DNA encoding novel
c 12		16.8	84.0	3227	22	AAH15736	Human cDNA sequenc
c 13		16.8	84.0	3242	22	AAH42470	Nucleotide sequenc
c 14		16.8	84.0	3284	22	AAH45368	Human DRF-1 coding
c 15		16.8	84.0	7686	21	AAZ29335	Human acrosome rea
c 16		16.8	84.0	109519	22	AAS08693	Micromonospora DNA
c 17		16.4	82.0	192	21	AAC07759	Human secreted pro
18		16.4	82.0	1442	16	AAT35111	Enhanced expressio
19		16.4	82.0	1467	20	AAV82457	Triticum sp. cyste
c 20		16.4	82.0	2025	21	AAA08830	Murine I kappa B p
c 21		16.4	82.0	2238	20	AAX98272	cDNA encoding huma
c 22		16.4	82.0	2238	20	AAX79305	Human IKK-alpha co
c 23		16.4	82.0	2238	20	AAX08918	Human IKK-alpha co
c 24		16.4	82.0	2238	20	AAV84690	Human IKK-alpha po
c 25		16.4	82.0	2238	21	AAC81424	Human I-kappa-B ki
c 26		16.4	82.0	2251	19	AAV22841	DNA encoding inhib
c 27		16.4	82.0	2272	19	AAV32969	Human I-kappa-B ki
c 28		16.4	82.0	2273	20	AAZ22536	DNA encoding a hum
c 29		16.4	82.0	2273	21	AAC81423	Human I-kappa-B ki
c 30		16.4	82.0	2991	19	AAV71078	IkappaB kinase alp
c 31		16.4	82.0	2994	19	AAV71077	Green fluorescent
32		16.4	82.0	3353	17	AAT03572	Human protocadheri
c 33		16.4	82.0	3466	24	ABI99798	Mouse ischaemic co
c 34		16.4	82.0	3579	21	AAF21147	Human low adenosin
c 35		16.4	82.0	3579	21	AAA35025	Human adenosine re
c 36		16.4	82.0	8631	21	AAF21150	Human low adenosin
c 37		16.4	82.0	8631	21	AAA35028	Human adenosine re
c 38		15.8	79.0	214	21	AAC09057	Human secreted pro
39		15.8	79.0	363	22	AAK56417	Human immune/haema
40		15.8	79.0	540	22	AAK67982	Human immune/haema
c 41		15.8	79.0	591	23	AAS92458	DNA encoding novel
c 42		15.8	79.0	662	22	AAH05070	Human cDNA clone (
c 43		15.8	79.0	727	22	AAP94066	Primer specific fo
c 44		15.8	79.0	744	22	AAH65417	C glutamicum codin
c 45		15.8	79.0	858	20	AAZ11393	M. vaccae antigen

ALIGNMENTS

RESULT 1
AAV17816
ID AAV17816 standard; cDNA; 2268 BP.
XX
AC AAV17816;
XX
DT 17-AUG-1998 (first entry)
XX
DE Homo sapiens don-1 gene splice variant.
XX
KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma;
KW epithelial cell; proliferation; stimulation; treatment; tumours;
KW skin; oesophagus; lung; breast; liver; pancreas; colon; prostate;
KW gastrointestinal tract; uterus; wound healing; transmembrane; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 69..2012
FT /*tag= a
FT /note= "don-1 polypeptide"
XX
PN WO9807736-A1.
XX
PD 26-FEB-1998.
XX
PF 18-AUG-1997; 97WO-US14585.
XX
PR 19-NOV-1996; 96US-0753007.
PR 19-AUG-1996; 96US-0699591.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

```
XX
PI   Busfield SJ,  Gearing DP;
XX
DR   WPI; 1998-169084/15.
DR   P-PSDB; AAW48383.
XX
PT   Mouse and human don-1 polypeptide(s) - useful for treatment of
PT   melanomas and adenocarcinoma(s), and for wound healing
XX
PS   Claim 4; Fig 7; 121pp; English.
XX
CC   The sequence is that of a human don-1 gene splice variant.
CC   Don-1 polypeptides stimulate proliferation of epithelial cells
CC   and thus are implicated in melanomas and adenocarcinomas in which
CC   epithelial cells proliferate out of control. Compounds that
CC   interfere with don-1 mediated cell proliferation can be used
CC   in the treatment of tumours such as melanomas and adenocarcinomas
CC   of the skin, oesophagus, lung, breast, liver, pancreas,
CC   gastrointestinal tract, colon, prostate or uterus. Alternatively,
CC   don-1 polypeptides can be used to stimulate epithelial cell
CC   proliferation, e.g. for wound healing.
XX
SQ   Sequence 2268 BP; 502 A; 735 C; 700 G; 331 T; 0 other;


  Query Match             87.0%;  Score 17.4;  DB 19;  Length 2268;
  Best Local Similarity   94.7%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||| ||||||||
Db   1532 cgcgctcggcggcagcctg 1550


RESULT   2
AAF81367/c
ID   AAF81367 standard; DNA; 7335 BP.
XX
AC   AAF81367;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Quorum sensing controlled gene qsc109 ORF.
XX
KW   Quorum sensing; antibacterial; bacterial signalling;
KW   opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
KW   immunosuppressive therapy; AIDS; ss.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200118248-A2.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-US24141.
XX
PR   03-SEP-1999;   99US-0153022.
XX
PA   (IOWA ) UNIV IOWA RES FOUND.
PA   (QUOR-) QUORUM SCI INC.
XX
PI   Whiteley M,  Lee KM,  Greenberg EP,  Muh U;
XX
DR   WPI; 2001-265973/27.
XX
PT   Identifying modulators of quorum sensing signaling in Pseudomonas
PT   aeruginosa bacteria, useful for treating infections in
PT   immunocompromized patients -
XX
PS   Claim 46; Page 108-110; 115pp; English.
XX
CC   Bacteria signal to one another to coordinate expression of specific genes
CC   in a cell density dependent fashion. This "bacterial signalling" is
CC   called "quorum sensing and response". Quorum sensing allows a bacterial
CC   species to sense its own number and regulate gene expression according to
CC   population density. The present sequence is an open reading frame (ORF)
CC   of a Pseudomonas aeruginosa quorum sensing controlled gene. Inhibitors of
CC   quorum sensing signalling renders a bacterial population more susceptible
CC   to treatment. The present invention relates to a method for identifying
CC   modulators of quorum sensing signalling in Pseudomonas aeruginosa
CC   bacteria. Modulators of quorum signaling may be used to treat P.
CC   aeruginosa infections. P. aeruginosa is an opportunistic pathogen of
CC   immunocompromised individuals (burn patients, cystic fibrosis patients,
CC   patients undergoing immunosuppressive therapy and patients with AIDS).
XX
SQ   Sequence 7335 BP; 1184 A; 2256 C; 2625 G; 1270 T; 0 other;


  Query Match             87.0%;  Score 17.4;  DB 22;  Length 7335;
  Best Local Similarity   94.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||||||||||| |
Db   4450 CGCGCTCGGCCGCAGCCAG 4432


RESULT   3
AAT37329
ID   AAT37329 standard; DNA; 28804 BP.
XX
AC   AAT37329;
XX
DT   30-NOV-1996  (first entry)
XX
DE   Sphingan biosynthetic gene region.
XX
KW   Sphingan; polysaccharide; spsB gene; glucosyl-IP-transferase; ds.
XX
OS   Sphinomonas strain S88 (ATCC 31554).
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1942..1944)
FT                   /*tag=  a
FT                   /codon_start= 1942..1944
FT                   /note= "spsG gene putative initiation codon"
FT   CDS             complement (3311..3313)
FT                   /*tag=  b
FT                   /codon_start= 3311..3313
FT                   /note= "spsS gene putative initiation codon"
FT   CDS             complement (5323..5325)
FT                   /*tag=  c
FT                   /codon_start= 5323..5325
FT                   /note= "spsR gene putative initiation codon"
FT   CDS             5526..5528
FT                   /*tag=  d
FT                   /codon_start= 5526..5528
FT                   /note= "spsQ gene putative initiation codon"
FT   CDS             complement (7076..7078)
FT                   /*tag=  e
FT                   /codon_start= 7076..7078
FT                   /note= "spsI gene putative initiation codon"
FT   CDS             7588..7590
FT                   /*tag=  f
FT                   /codon_start= 7588..7590
FT                   /note= "spsK gene putative initiation codon"
FT   CDS             8643..8645
FT                   /*tag=  g
FT                   /codon_start= 8643..8645
FT                   /note= "spsL gene putative initiation codon"
FT   CDS             complement (10938..10940)
FT                   /*tag=  h
FT                   /codon_start= 10938..10940
FT                   /note= "spsJ gene putative initiation codon"
FT   CDS             11569..11571
FT                   /*tag=  i
```

```
FT                   /codon_start= 11569..11571
FT                   /note= "spsF gene putative initiation codon"
FT   CDS             12886..12888
FT                   /*tag=  j
FT                   /codon_start= 12886..12888
FT                   /note= "spsD gene putative initiation codon"
FT   CDS             15165..15167
FT                   /*tag=  k
FT                   /codon_start= 15165..15167
FT                   /note= "spsE gene putative initiation codon"
FT   CDS             15883..15885
FT                   /*tag=  l
FT                   /codon_start= 15883..15885
FT                   /note= "urf32 gene putative initiation codon"
FT   CDS             16748..16750
FT                   /*tag=  m
FT                   /codon_start= 16748..16750
FT                   /note= "urf26 gene putative initiation codon"
FT   CDS             complement (21082..21084)
FT                   /*tag=  n
FT                   /codon_start= 21082..21084
FT                   /note= "atrB gene putative initiation codon"
FT   CDS             21706..23118
FT                   /*tag=  o
FT                   /label= spsB
FT                   /product= glucosyl IP-transferase
FT   CDS             23238..23240
FT                   /*tag=  p
FT                   /codon_start= 23238..23240
FT                   /note= "rhsA gene putative initiation codon"
FT   CDS             24113..24115
FT                   /*tag=  q
FT                   /codon_start= 24113..24115
FT                   /note= "rhsC gene putative initiation codon"
FT   CDS             24683..24685
FT                   /*tag=  r
FT                   /codon_start= 24683..24685
FT                   /note= "rhsB gene putative initiation codon"
FT   CDS             25744..25746
FT                   /*tag=  s
FT                   /codon_start= 25744..25746
FT                   /note= "rhsD gene putative initiation codon"
FT   CDS             complement (27534..27536)
FT                   /*tag=  t
FT                   /codon_start= 27534..27536
FT                   /note= "urf31 gene putative initiation codon"
FT   CDS             27747
FT                   /*tag=  u
FT                   /codon_start= 27747..27749
FT                   /note= "urf34 gene putative initiation codon"
XX
PN   EP728841-A2.
XX
PD   28-AUG-1996.
XX
PF   24-JAN-1996;   96EP-0300467.
XX
PR   24-JAN-1995;   95US-0377440.
XX
PA   (SHIN-) SHINETSU BIO INC.
PA   (SHIE ) SHINETSU CHEM CO LTD.
XX
PI   Armentrout RW,  Mikolajczak M,  Pollock TJ,  Thorne L;
PI   Yamazaki M;
XX
DR   WPI; 1996-386292/39.
DR   P-PSDB; AAW03997.
XX
PT   New isolated DNA from Sphingomonas sp. - used for transforming
PT   recipient bacteria to obtain hyper-producers of sphingan
PT   polysaccharide(s).
XX
PS   Claim 32; Page 56-70; 105pp; English.
XX
CC   A 28.8 kb chromosomal fragment of Sphingomonas strain S88 was
CC   isolated on the basis of its ability to restore sphingan
CC   biosynthetic capability to Sphingomonas mutant S88m260.  It
CC   contains 23-25 genes, including sps genes coding for biosynthesis of
CC   the polysaccharide sphingan, rhs genes coding for dTDP-(L)rhamnose
CC   biosynthesis, atrDB genes coding for a transport function and some
CC   unidentified open translation reading frames (urf).  The spsB gene
CC   was identified that is believed to code for glucosyl IP-transferase
CC   (AAW03997), an enzyme catalysing the first step of assembly of
CC   sphingan carbohydrates.  DNA fragments of S88 can be inserted into
CC   a vector in multiple copies and used to produce engineered bacteria
CC   that are hyper-producers of sphingan.
XX
SQ   Sequence 28804 BP; 4974 A; 9806 C; 9228 G; 4796 T; 0 other;


  Query Match             87.0%;  Score 17.4;  DB 17;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcct 19
           ||||||| ||||||||||| 
Db    7745 ccgcgctgggccgcagcct 7763


RESULT   4
AAT92474
ID   AAT92474 standard; DNA; 28804 BP.
XX
AC   AAT92474;
XX
DT   04-FEB-1998  (first entry)
XX
DE   Sphingomonas genus microbe isolated DNA sequence producing sphingan.
XX
KW   Sphingomonas microbe; sphingan polysaccharide biosynthesis gene;
KW   sphingan S-88; spsB gene; ss.
XX
OS   Sphingomonas sp.
XX
PN   JP09252775-A.
XX
PD   30-SEP-1997.
XX
PF   24-JAN-1996;   96JP-0043977.
XX
PR   16-JAN-1996;   96JP-0004621.
PR   24-JAN-1995;   95US-0377440.
XX
PA   (SHIN-) SHINETSU BIO INC.
PA   (SHIE ) SHINETSU CHEM IND CO LTD.
XX
DR   WPI; 1997-530148/49.
XX
PT   Sphingan polysaccharide biosynthesis gene from Sphingomonas microbe
PT   - can be introduced into Sphingomonas host cells to increase
PT   sphingan production
XX
PS   Claim 74; Page 34-43; 55pp; Japanese.
XX
CC   The present sequence represents a new DNA sequence which has been
CC   isolated from a sphingan polysaccharide-producing microbe of the
CC   genus Sphingomonas. When recombined in a Sphingomonas host, the
CC   presence of several copies of the sphingan polysaccharide biosynthesis
CC   gene results in a microbe which expresses high levels of sphingan.
CC   A method has been developed for producing a Sphingomonas microbe with
CC   enhanced sphingan polysaccharide expression, by transforming the cell
CC   with the sphingan polysaccharide biosynthesis gene.  The method can
CC   produce sphingan in large quantities.
XX
SQ   Sequence 28804 BP; 4974 A; 9804 C; 9230 G; 4796 T; 0 other;
```

```
  Query Match             87.0%;  Score 17.4;  DB 18;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||
Db   7745 ccgcgctgggccgcagcct 7763


RESULT   5
AAV99812
ID   AAV99812 standard; DNA; 28804 BP.
XX
AC   AAV99812;
XX
DT   14-JUN-1999  (first entry)
XX
DE   Sphingomonas S88 sps gene cluster.
XX
KW   Xantham gum; gum; exopolysaccharide; gum gene cluster;
KW   recombination; Xanthomonas; Sphingomonas; sphingan S88; sugar;
KW   substrate; lactose; sucrose; starch; ss.
XX
OS   Sphingomonas sp. S88.
XX
PN   WO9856942-A1.
XX
PD   17-DEC-1998.
XX
PF   12-JUN-1998;   98WO-US12322.
XX
PR   12-JUN-1998;   98US-0096942.
PR   12-JUN-1997;   97US-0049428.
PR   11-JUN-1998;   98US-0096867.
XX
PA   (SHIN-) SHINETSU BIO INC.
PA   (SHIE ) SHINETSU CHEM CO LTD.
XX
PI   Armentrout RW,  Mikolajczak M,  Pollock TJ,  Thorne L;
PI   Yamazaki M;
XX
DR   WPI; 1999-080915/07.
XX
PT   Production of exopolysaccharide, e.g. xanthan gum by a Sphingomonas
PT   species bacterium - containing the gumB-M genes of X. campestris,
PT   and using lactose as a C source
XX
PS   Example 1; Page 36-46; 50pp; English.
XX
CC   Xanthan gum is an exopolysaccharide produced by Xanthomonas species
CC   and encoded by the gum gene cluster.  The gum gene cluster of
CC   Xanthomonas campestris can be inserted into a Sphingomonas species
CC   of bacterium.  The gum gene cluster is inserted into the Sphingomonas
CC   S88 sps gene cluster which itself expresses an exopolysaccharide,
CC   sphingan S88. The new novel recombinant Sphingomonas strain can then
CC   produce xanthan gum from sugar substrates i.e. the cheese-making
CC   by-product whey lactose, sucrose or starch. These are substrates
CC   which the Xanthomonas campestris bacterium cannot utilise in its
CC   production of Xantham gum. Use of the Sphingomonas species also
CC   minimises the presence of contaminating cellulases in the growth
CC   medium into which the xantham gum is secreted.
XX
SQ   Sequence 28804 BP; 4974 A; 9804 C; 9230 G; 4796 T; 0 other;


  Query Match             87.0%;  Score 17.4;  DB 20;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||
Db   7745 ccgcgctgggccgcagcct 7763


RESULT   6
AAV81474
ID   AAV81474 standard; DNA; 28804 BP.
XX
AC   AAV81474;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Chromosomal fragment of Sphingomonas strain 88 genome.
XX
KW   Chromosome; sphingan; bacterium; polysaccharide; polymer; additive; food;
KW   glycosyl-C55-isoprenylphosphate transferase; textile; cosmetic; paper;
KW   paint; cement; viscosity; adhesive; petroleum; chemical; ds.
XX
OS   Sphingomonas sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             21706..23118
FT                   /*tag= a
FT                   /product= "spsB protein"
XX
PN   US5854034-A.
XX
PD   29-DEC-1998.
XX
PF   24-JAN-1996;   96US-0592874.
XX
PR   24-JAN-1996;   96US-0592874.
PR   24-JAN-1995;   95US-0377440.
XX
PA   (SHIN-) SHIN-ETSU CO INC.
PA   (SHIE ) SHINETSU CHEM CO LTD.
XX
PI   Armentrout RW,  Mikolajczak M,  Pollock TJ,  Thorne L;
PI   Yamazaki M;
XX
DR   WPI; 1999-094909/08.
DR   P-PSDB; AAW67750.
XX
PT   Production of sphingan polysaccharide products - by introducing DNA
PT   from sphingan-producing Sphingomonas species in multiple copies into
PT   recipient Sphingomonas sp.
XX
PS   Example 16; Fig 14A-K; 66pp; English.
XX
CC   This sequence represents a chromosomal fragment of the Sphingomonas sp.
CC   strain S88 genome.  DNA from this sphingan-producing bacterium can be
CC   used to increase production of sphingan polysaccharides in other
CC   microorganisms, especially other Sphingomonas strains.  The DNA
CC   transferred to other strains includes the spsB gene which encodes a
CC   glycosyl-C55-isoprenylphosphate (glycosyl-IP) transferase.  The sphingans
CC   are useful as speciality polymers and as additives in textile
CC   applications, foods, cosmetics, paper, paint, cements, e.g. as viscosity
CC   modifiers, in various other coating applications, and as adhesives and
CC   additives to petroleum products and speciality chemicals.
XX
SQ   Sequence 28804 BP; 4974 A; 9804 C; 9230 G; 4796 T; 0 other;


  Query Match             87.0%;  Score 17.4;  DB 20;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||
Db   7745 ccgcgctgggccgcagcct 7763


RESULT   7
AAF65897/c
ID   AAF65897 standard; cDNA; 409 BP.
```

```
XX
AC   AAF65897;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1653.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 781; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 409 BP; 108 A; 80 C; 121 G; 100 T; 0 other;

  Query Match             84.0%;  Score 16.8;  DB 22;  Length 409;
  Best Local Similarity   90.0%;  Pred. No. 3.7e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcctg 20
           ||||||||||| |||||||
Db      24 CCGCGCTCGGCTACAGCCTG 5


RESULT   8
AAH05140/c
ID   AAH05140 standard; cDNA; 734 BP.
XX
AC   AAH05140;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:1975.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 1975; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 734 BP; 205 A; 139 C; 183 G; 194 T; 13 other;

  Query Match             84.0%;  Score 16.8;  DB 22;  Length 734;
  Best Local Similarity   90.0%;  Pred. No. 3.5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcctg 20
           ||||||||||| |||||||
Db      64 CCGCGCTCGGCTACAGCCTG 45


RESULT   9
AAV12277
ID   AAV12277 standard; DNA; 819 BP.
XX
AC   AAV12277;
XX
```

```
DT   12-MAY-1998  (first entry)
XX
DE   Maleate cis-trans isomerase encoding DNA.
XX
KW   Maleate cis-trans isomerase; maleic acid; fumaric acid; antioxidant;
KW   baking powder; fruit juice; tartaric acid; citric acid; ds.
XX
OS   Arthrobacter globiformis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..819
FT                   /*tag= a
FT                   /product= "maleate cis-trans-isomerase"
XX
PN   JP10033181-A.
XX
PD   10-FEB-1998.
XX
PF   24-JUL-1996;   96JP-0194961.
XX
PR   24-JUL-1996;   96JP-0194961.
XX
PA   (MITU ) MITSUBISHI CHEM CORP.
XX
DR   WPI; 1998-172098/16.
DR   P-PSDB; AAW44186.
XX
PT   Maleate-cis-trans-isomerase from Arthrobacter globiformis IAM 12102
PT   - used for the efficient production of fumaric acid by isomerisation
PT   of maleic aid
XX
PS   Claim 1; Page 10-11; 14pp; Japanese.
XX
CC   The present sequence encodes maleate cis-trans isomerase, isolated
CC   from Arthrobacter globiformis IAM 12102. Maleate cis-trans isomerase
CC   has the activity to form fumaric acid by isomerisation of maleic acid.
CC   The enzyme or transformant producing the enzyme can be added to an
CC   aqueous solution containing maleic acid to produce fumaric acid.
CC   Fumaric acid is useful as a substitute for tartaric acid in beverages
CC   and baking powders and as a replacement for citric acid in fruit
CC   drinks. Fumaric acid is also used as an antioxidant.
XX
SQ   Sequence 819 BP; 155 A; 299 C; 250 G; 115 T; 0 other;


  Query Match             84.0%;  Score 16.8;  DB 19;  Length 819;
  Best Local Similarity   90.0%;  Pred. No. 3.5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcctg 20
           ||||||| | ||||||||||
Db     581 ccgcgcttgcccgcagcctg 600


RESULT  10
AAH46831/c
ID   AAH46831 standard; cDNA; 1877 BP.
XX
AC   AAH46831;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human immunoglobulin domain-containing polypeptide encoding cDNA.
XX
KW   Immunoglobulin domain-containing polypeptide; secreted protein; human;
KW   cytostatic; immunosuppressive; antiinflammatory; antirheumatic; vaccine;
KW   antiarthritic; osteopathic; antipsoriatic; antibacterial; fungicide;
KW   nootropic; virucide; neuroprotective; antiparkinsonian; antiallergic;
KW   antiasthmatic; dermatological; cerebroprotective; antidiabetic;
KW   hemostatic; gene therapy; ss.
XX
OS   Homo sapiens.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             54..920
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            54..363
FT                   /*tag= b
FT                   /number= "1"
FT   intron          364..621
FT                   /*tag= c
FT                   /number= "1"
FT   exon            622..917
FT                   /*tag= d
FT                   /number= "2"
XX
PN   WO200155334-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02461.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   04-OCT-2000; 2000US-0680849.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Boyle BJ,  Arterburn MC,  Tang YT,  Liu C,  Drmanac RA;
XX
DR   WPI; 2001-476197/51.
DR   P-PSDB; AAB85464.
XX
PT   Novel immunoglobulin domain-containing polypeptides and polynucleotides
PT   useful for diagnosis, prevention, treatment of cancer, neurological
PT   disorders, immunological diseases, such as psoriasis, rheumatoid
PT   arthritis -
XX
PS   Claim 1; Page 125-127; 132pp; English.
XX
CC   The invention provides novel human secreted immunoglobulin domain-
CC   containing polypeptides (I) and polynucleotides (II) encoding them.
CC   The polypeptide exhibits cytokine, cell proliferation or cell
CC   differentiation, stem cell growth factor activity and activin or inhibin
CC   related activities, chemotactic or chemokinetic activities. (I) is useful
CC   for re-engineering damaged or diseased tissues, transplantation,
CC   manufacture of bio-pharmaceuticals and development of bio-sensors (see
CC   AAH46827 for a detailed description of the uses and diseases that can be
CC   treated by using the polypeptides of the invention). The present sequence
CC   represents a immunoglobulin domain-containing polypeptide encoding cDNA.
XX
SQ   Sequence 1877 BP; 518 A; 398 C; 471 G; 489 T; 1 other;


  Query Match             84.0%;  Score 16.8;  DB 22;  Length 1877;
  Best Local Similarity   90.0%;  Pred. No. 3.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcctg 20
           |||||| |||||||| ||||
Db    1058 CCGCGCCCGGCCGCAACCTG 1039


RESULT  11
AAS84820/c
ID   AAS84820 standard; cDNA; 2934 BP.
XX
AC   AAS84820;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20624.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG20633.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 20624; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2934 BP; 947 A; 518 C; 665 G; 804 T; 0 other;

Query Match 84.0%; Score 16.8; DB 23; Length 2934;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgcgctcggccgcagcctg 20
||||||||||| |||||||
Db 51 CCGCGCTCGGCTACAGCCTG 32

RESULT 12
AAH15736/c
ID AAH15736 standard; cDNA; 3227 BP.
XX
AC AAH15736;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14141.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14141; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 3227 BP; 1047 A; 570 C; 728 G; 882 T; 0 other;

Query Match 84.0%; Score 16.8; DB 22; Length 3227;
Best Local Similarity 90.0%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgcgctcggccgcagcctg 20
||||||||||| |||||||
Db 65 CCGCGCTCGGCTACAGCCTG 46

RESULT 13
AAH42470/c
ID AAH42470 standard; DNA; 3242 BP.
XX
AC AAH42470;
XX
DT 01-OCT-2001 (first entry)
XX
DE Nucleotide sequence of human cancer associated antigen MEL-2.
XX

```
KW   Cancer associated antigen; MEL-1; MEL-2; vaccine; cancer; marker; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155389-A1.
XX
PD   02-AUG-2001.
XX
PF   03-JAN-2001; 2001WO-US00153.
XX
PR   28-JAN-2000; 2000US-0493914.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (CORR ) CORNELL RES FOUND INC.
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Jaeger D,  Stockert E,  Scanlan M,  Gure A,  Jaeger E,  Knuth A;
PI   Lloyd OJ,  Chen Y;
XX
DR   WPI; 2001-465574/50.
XX
PT   Novel cancer associated antigen useful as a vaccine for treating a
PT   cancerous condition and as a marker for the diagnosis, determination of
PT   regression or progression of onset of a cancerous condition  -
XX
PS   Claim 6; Page 28-30; 36pp; English.
XX
CC   The present sequence encodes a human cancer associated antigen,
CC   designated  MEL-2. The specification also describes MEL-1. The
CC   cancer associated antigens are useful in immunogenic compositions
CC   or vaccines for treating a cancerous condition. They may also be
CC   used as markers for the diagnosis of cancer and for determination
CC   of regression or progression of onset of a cancerous condition.
XX
SQ   Sequence 3242 BP; 1082 A; 559 C; 714 G; 878 T; 9 other;


  Query Match             84.0%;  Score 16.8;  DB 22;  Length 3242;
  Best Local Similarity   90.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcctg 20
          ||||||||||| |||||||
Db     37 CCGCGCTCGGCTACAGCCTG 18


RESULT  14
AAH45368/c
ID   AAH45368 standard; cDNA; 3284 BP.
XX
AC   AAH45368;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Human DRF-1 coding sequence.
XX
KW   Human; differentiation responsive factor-1; DRF-1; haemostatic;
KW   cell differentiation; haematopoietic stem cell; megakaryocyte;
KW   thrombocytopaenia; ss.
XX
OS   Homo sapiens.
XX
PN   JP2001122898-A.
XX
PD   08-MAY-2001.
XX
PF   25-OCT-1999;   99JP-0302700.
XX
PR   25-OCT-1999;   99JP-0302700.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2001-364766/38.
DR   P-PSDB; AAG64040.
XX
PT   A differentiation responsive factor-1 (DRF-1) and a gene encoding for
PT   DRF-1.  -
XX
PS   Disclosure; Fig 3; 13pp; Japanese.
XX
CC   The present sequence encodes differentiation responsive factor-1 (DRF-1).
CC   DRF-1 is capable of inducing differentiation of haematopoietic cells,
CC   particularly haematopoietic stem cells, to megakaryocytes in
CC   conjunction with a chemical substance such as DMSO, TPO or K252a.
CC   The DRF-1 polynucleotide and polypeptide are useful for the prevention
CC   and treatment of diseases associated with thrombocytopaenia.
XX
SQ   Sequence 3284 BP; 1073 A; 578 C; 738 G; 895 T; 0 other;


  Query Match             84.0%;  Score 16.8;  DB 22;  Length 3284;
  Best Local Similarity   90.0%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcctg 20
          ||||||||||| |||||||
Db    108 CCGCGCTCGGCTACAGCCTG 89


RESULT  15
AAZ29335/c
ID   AAZ29335 standard; DNA; 7686 BP.
XX
AC   AAZ29335;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Human acrosome reaction protein-PKDREJ encoding DNA.
XX
KW   Acrosome reaction protein; P3; sperm protein; polycystin-1; suREJ;
KW   Polycystic Kidney Disease and Receptor for Egg Jelly protein; PKDREJ;
KW   zona pellucida; fertility; contraceptive; gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..6762
FT                   /*tag=  a
FT                   /product= "Acrosome reaction protein (PKDREJ)"
FT   polyA_signal    7636..7641
FT                   /*tag=  b
XX
PN   WO9964457-A1.
XX
PD   16-DEC-1999.
XX
PF   10-JUN-1999;   99WO-GB01839.
XX
PR   10-JUN-1998;   98GB-0012534.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Harris PC,  Hugues JR,  Ward CJ;
XX
DR   WPI; 2000-097518/08.
DR   P-PSDB; AAY44301.
XX
PT   New functional mammalian acrosome reaction protein, useful for
PT   fertility treatment  -
XX
PS   Claim 8; Fig 3; 40pp; English.
XX
CC   The present sequence is a DNA encoding functional human acrosome
CC   reaction protein, PKDREJ or P3. PKDREJ is a sperm protein involved in
CC   binding the sperm to the egg and/or triggering the acrosome reaction.
CC   It exhibits homology to human polycystin-1 and suREJ. It can be used to
```

CC identify antibodies or proteins that block the acrosome reaction and
CC compounds that trigger acrosome reaction in the absence of zona
CC pellucida. Molecules that enhance the efficacy of the acrosome reaction
CC protein can be used to increase fertility and those blocking its
CC action can be used as contraceptives. DNA encoding PKDREJ can be used in
CC gene therapy and also as primer or probe for identifying sequences that
CC encode mutant forms of acrosome reaction protein.
XX
SQ Sequence 7686 BP; 1952 A; 1827 C; 1809 G; 2098 T; 0 other;

Query Match 84.0%; Score 16.8; DB 21; Length 7686;
Best Local Similarity 90.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccgcgctcggccgcagcctg 20
||||||| |||||||| |||
Db 190 CCGCGCTGGGCCGCAGACTG 171

Search completed: June 23, 2002, 14:59:54
Job time: 81126 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 22, 2002, 17:19:36 ; Search time 167.81 Seconds
(without alignments)
29.275 Million cell updates/sec

Title: US-09-747-514A-1
Perfect score: 20
Sequence: 1 ccgcgctcggccgcagcctg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.4	87.0	2268	3	US-08-753-007A-31	Sequence 31, Appl
	2	17.4	87.0	2268	3	US-09-398-496-31	Sequence 31, Appl
c	3	17.4	87.0	15297	4	US-09-817-180-3	Sequence 3, Appli
	4	17.4	87.0	23673	4	US-09-773-816-1	Sequence 1, Appli
	5	17.4	87.0	28804	2	US-08-592-874-1	Sequence 1, Appli
	6	17.4	87.0	28804	3	US-09-096-942-2	Sequence 2, Appli
	7	17.4	87.0	28804	3	US-09-096-867-2	Sequence 2, Appli
c	8	16.4	82.0	2025	3	US-09-160-483-2	Sequence 2, Appli
c	9	16.4	82.0	2238	2	US-08-890-853-3	Sequence 3, Appli
c	10	16.4	82.0	2238	2	US-09-099-125A-3	Sequence 3, Appli
c	11	16.4	82.0	2238	2	US-09-099-124A-3	Sequence 3, Appli
c	12	16.4	82.0	2238	4	US-09-032-476-3	Sequence 3, Appli
c	13	16.4	82.0	2238	4	US-08-890-854-3	Sequence 3, Appli
c	14	16.4	82.0	2238	4	US-09-023-324-3	Sequence 3, Appli
c	15	16.4	82.0	2251	4	US-08-910-820-7	Sequence 7, Appli
c	16	16.4	82.0	2273	2	US-09-197-360-1	Sequence 1, Appli
c	17	16.4	82.0	2273	4	US-09-168-629-1	Sequence 1, Appli
c	18	16.4	82.0	2273	4	US-08-810-131A-1	Sequence 1, Appli
	19	16.4	82.0	3353	1	US-08-453-695A-109	Sequence 109, App
	20	16.4	82.0	3353	1	US-08-268-161A-109	Sequence 109, App
	21	16.4	82.0	3353	2	US-08-453-702A-109	Sequence 109, App
	22	16.4	82.0	3353	4	US-09-099-639-109	Sequence 109, App
	23	16.4	82.0	3353	5	PCT-US95-08071-109	Sequence 109, App
c	24	15.8	79.0	858	4	US-09-095-855-200	Sequence 200, App
c	25	15.8	79.0	1875	4	US-09-422-869-21	Sequence 21, Appl
c	26	15.8	79.0	30001	1	US-08-125-468-1	Sequence 1, Appli
c	27	15.8	79.0	30001	2	US-08-474-933-1	Sequence 1, Appli
c	28	15.8	79.0	49136	4	US-09-422-869-1	Sequence 1, Appli
c	29	15.4	77.0	649	4	US-08-998-416-156	Sequence 156, App
c	30	15.4	77.0	1364	4	US-09-095-855-204	Sequence 204, App
c	31	15.4	77.0	1664	3	US-09-339-993-1	Sequence 1, Appli
c	32	15.2	76.0	1089	3	US-09-195-666A-10	Sequence 10, Appl
c	33	15.2	76.0	1272	4	US-09-191-136-13	Sequence 13, Appl
c	34	15.2	76.0	1695	3	US-09-008-481A-3	Sequence 3, Appli
c	35	15.2	76.0	1695	3	US-09-195-666A-17	Sequence 17, Appl
c	36	15.2	76.0	1695	4	US-09-309-592-3	Sequence 3, Appli
c	37	15.2	76.0	1928	3	US-09-008-481A-9	Sequence 9, Appli
c	38	15.2	76.0	1928	3	US-09-195-666A-15	Sequence 15, Appl
c	39	15.2	76.0	1928	4	US-09-309-592-9	Sequence 9, Appli
	40	15.2	76.0	1939	6	5198542-3	Patent No. 5198542
	41	15.2	76.0	2839	1	US-07-814-964-6	Sequence 6, Appli
	42	15.2	76.0	2839	1	US-08-258-442-6	Sequence 6, Appli
	43	15.2	76.0	2839	1	US-08-328-809-1	Sequence 1, Appli
	44	15.2	76.0	2839	3	US-09-015-003-1	Sequence 1, Appli
	45	15.2	76.0	2839	5	PCT-US92-11107-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-08-753-007A-31
; Sequence 31, Application US/08753007A
; Patent No. 6074841
;  GENERAL INFORMATION:
;    APPLICANT:  Gearing, David P.
;    APPLICANT:  Busfield, Samantha J.
;    TITLE OF INVENTION:  DON-1 GENE AND POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/753,007A
;      FILING DATE:  19-NOV-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/699,591
;      FILING DATE:  19-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07334/022001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  69...2009
;      OTHER INFORMATION:
US-08-753-007A-31
```

```
  Query Match             87.0%;  Score 17.4;  DB 3;  Length 2268;
  Best Local Similarity   94.7%;  Pred. No. 43;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||| ||||||||
Db   1532 CGCGCTCGGCGGCAGCCTG 1550


RESULT   2
US-09-398-496-31
; Sequence 31, Application US/09398496
; Patent No. 6133423
;  GENERAL INFORMATION:
;    APPLICANT:  Gearing, David P.
;    APPLICANT:  Busfield, Samantha J.
;    TITLE OF INVENTION:  DON-1 GENE AND POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/398,496
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/753,007
;      FILING DATE:  19-NOV-1996
;      APPLICATION NUMBER:  08/699,591
;      FILING DATE:  19-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07334/022001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  69...2009
;      OTHER INFORMATION:
US-09-398-496-31


  Query Match             87.0%;  Score 17.4;  DB 3;  Length 2268;
  Best Local Similarity   94.7%;  Pred. No. 43;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          |||||||||| ||||||||
Db   1532 CGCGCTCGGCGGCAGCCTG 1550


RESULT   3
```

```
US-09-817-180-3/c
; Sequence 3, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001183
;  CURRENT APPLICATION NUMBER: US/09/817,180
;  CURRENT FILING DATE:  2001-03-27
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 15297
;   TYPE: DNA
;   ORGANISM: Human
US-09-817-180-3


  Query Match             87.0%;  Score 17.4;  DB 4;  Length 15297;
  Best Local Similarity   94.7%;  Pred. No. 35;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||||
Db   1873 CCGCGCTGGGCCGCAGCCT 1855


RESULT   4
US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
;  APPLICANT: Stanford University
;  APPLICANT:  Khosla, Chaitan
;  TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
;  TITLE OF INVENTION:  ANTAGONISTS
;  FILE REFERENCE: 28600-20210.00
;  CURRENT APPLICATION NUMBER: US/09/773,816
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/243,458
;  PRIOR FILING DATE: 2000-10-25
;  PRIOR APPLICATION NUMBER: US 60/179,305
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 23673
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(23623)
;   OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1


  Query Match             87.0%;  Score 17.4;  DB 4;  Length 23673;
  Best Local Similarity   94.7%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgcgctcggccgcagcctg 20
          ||||||| ||||||||||||
Db   3324 cgcgctgggccgcagcctg 3342


RESULT   5
US-08-592-874-1
; Sequence 1, Application US/08592874
; Patent No. 5854034
;  GENERAL INFORMATION:
;    APPLICANT:  POLLOCK, THOMAS J.
```

```
;    APPLICANT:  YAMAZAKI, MOTOHIDE
;    APPLICANT:  THORNE, LINDA
;    APPLICANT:  MIKOLAJCZAK, MARCIA
;    APPLICANT:  ARMENTROUT, RICHARD W.
;    TITLE OF INVENTION:  DNA SEGMENTS AND METHODS FOR INCREASING
;    TITLE OF INVENTION:  POLYSACCHARIDE PRODUCTION
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JULES E. GOLDBERG
;      STREET:  261 MADISON AVENUE
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10016-2391
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,874
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/377,440
;      FILING DATE:  24-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GOLDBERG, JULES E.
;      REGISTRATION NUMBER:  24,408
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-986-4090
;      TELEFAX:  212-818-9479
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FRAGMENT TYPE:  N-terminal
US-08-592-874-1


  Query Match             87.0%;  Score 17.4;  DB 2;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||||
Db   7745 CCGCGCTGGGCCGCAGCCT 7763


RESULT   6
US-09-096-942-2
; Sequence 2, Application US/09096942
; Patent No. 6027925
; GENERAL INFORMATION:
;  APPLICANT: Pollock, Thomas J
;  APPLICANT: Mikolajczak, Marcia
;  APPLICANT: Yamazaki, Motohide
;  APPLICANT: Thorne, Linda
;  APPLICANT: Armentrout, Richard W
;  TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
;  TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
;  FILE REFERENCE: seq list for appl filed from pro. appl
;  CURRENT APPLICATION NUMBER: US/09/096,942
;  CURRENT FILING DATE: 1998-06-12
;  EARLIER APPLICATION NUMBER: 60/049,428
;  EARLIER FILING DATE: 1997-06-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 28804
;   TYPE: DNA
;   ORGANISM: Sphingomonas sp. S88
US-09-096-942-2


  Query Match             87.0%;  Score 17.4;  DB 3;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||||
Db   7745 ccgcgctgggccgcagcct 7763


RESULT   7
US-09-096-867-2
; Sequence 2, Application US/09096867
; Patent No. 6030817
; GENERAL INFORMATION:
;  APPLICANT: Pollock, Thomas J
;  APPLICANT: Mikolajczak, Marcia
;  APPLICANT: Yamazaki, Motohide
;  APPLICANT: Thorne, Linda
;  APPLICANT: Armentrout, Richard W
;  TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
;  TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
;  FILE REFERENCE: seq list for appl filed from pro. appl
;  CURRENT APPLICATION NUMBER: US/09/096,867
;  CURRENT FILING DATE: 1998-06-11
;  EARLIER APPLICATION NUMBER: 60/049,428
;  EARLIER FILING DATE: 1997-06-12
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 28804
;   TYPE: DNA
;   ORGANISM: Sphingomonas sp. S88
US-09-096-867-2


  Query Match             87.0%;  Score 17.4;  DB 3;  Length 28804;
  Best Local Similarity   94.7%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcct 19
          ||||||| ||||||||||||
Db   7745 ccgcgctgggccgcagcct 7763


RESULT   8
US-09-160-483-2/c
; Sequence 2, Application US/09160483A
; Patent No. 6083732
; GENERAL INFORMATION:
;  APPLICANT: Marcu, Kenneth B.
;  TITLE OF INVENTION: A  BIOLOGICALLY ACTIVE ALTERNATIVE FORM OF THE IKK
;  TITLE OF INVENTION: ALPHA IKB KINASE
;  FILE REFERENCE: Docket No. 6083732: 178-255
;  CURRENT APPLICATION NUMBER: US/09/160,483A
;  CURRENT FILING DATE: 1998-09-25
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2025
;   TYPE: DNA
;   ORGANISM: mus.musculus domesticus
US-09-160-483-2


  Query Match             82.0%;  Score 16.4;  DB 3;  Length 2025;
  Best Local Similarity   94.4%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 ccgcgctcggccgcagcc 18
           |||||| ||||||||||||
Db      87 CCGCGCCCGGCCGCAGCC 70


RESULT   9
US-08-890-853-3/c
; Sequence 3, Application US/08890853
; Patent No. 5851812
;  GENERAL INFORMATION:
;    APPLICANT:  Goeddel, David V.
;    APPLICANT:  Woronicz, John
;    TITLE OF INVENTION:  IKK-  Proteins, Nucleic Acids and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,853
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-006-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-890-853-3


  Query Match             82.0%;  Score 16.4;  DB 2;  Length 2238;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcc 18
           |||||| ||||||||||||
Db      34 CCGCGCCCGGCCGCAGCC 17


RESULT  10
US-09-099-125A-3/c
; Sequence 3, Application US/09099125A
; Patent No. 5916760
;  GENERAL INFORMATION:
;    APPLICANT:  Goeddel, David V.
;    APPLICANT:  Woronicz, John
;    TITLE OF INVENTION:  IKK-  Proteins, Nucleic Acids and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/099,125A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-006-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-099-125A-3


  Query Match             82.0%;  Score 16.4;  DB 2;  Length 2238;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcc 18
           |||||| ||||||||||||
Db      34 CCGCGCCCGGCCGCAGCC 17


RESULT  11
US-09-099-124A-3/c
; Sequence 3, Application US/09099124A
; Patent No. 5939302
;  GENERAL INFORMATION:
;    APPLICANT:  Goeddel, David V.
;    APPLICANT:  Woronicz, John
;    TITLE OF INVENTION:  IKK-  Proteins, Nucleic Acids and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/099,124A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-006-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-099-124A-3


  Query Match             82.0%;  Score 16.4;  DB 2;  Length 2238;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcc 18
          |||||| ||||||||||||
Db     34 CCGCGCCCGGCCGCAGCC 17


RESULT  12
US-09-032-476-3/c
; Sequence 3, Application US/09032476
; Patent No. 6235492
;  GENERAL INFORMATION:
;    APPLICANT:  Rothe, Mike
;    APPLICANT:  Cao, Zhaodan
;    APPLICANT:  R gnier, Catherine
;    TITLE OF INVENTION:  IKK-` Proteins, Nucleic Acids and Methods
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/032,476
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/890,854
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-006-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-032-476-3


  Query Match             82.0%;  Score 16.4;  DB 4;  Length 2238;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcc 18
          |||||| ||||||||||||
Db     34 CCGCGCCCGGCCGCAGCC 17


RESULT  13
US-08-890-854-3/c
; Sequence 3, Application US/08890854
; Patent No. 6235512
;  GENERAL INFORMATION:
;    APPLICANT:  Rothe, Mike
;    APPLICANT:  Cao, Zhaodan
;    APPLICANT:  R gnier, Catherine
;    TITLE OF INVENTION:  IKK-` Proteins, Nucleic Acids and Methods
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/890,854
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-006-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-890-854-3


  Query Match             82.0%;  Score 16.4;  DB 4;  Length 2238;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ccgcgctcggccgcagcc 18
          |||||| ||||||||||||
Db     34 CCGCGCCCGGCCGCAGCC 17


RESULT  14
US-09-023-324-3/c
; Sequence 3, Application US/09023324
; Patent No. 6235513
;  GENERAL INFORMATION:
;    APPLICANT:  Rothe, Mike
;    APPLICANT:  Cao, Zhaodan
;    APPLICANT:  R gnier, Catherine
;    TITLE OF INVENTION:  IKK-` Proteins, Nucleic Acids and Methods
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/023,324
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/890,854
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  OSMAN, RICHARD A
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  T97-006-1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 343-4341
;       TELEFAX:  (415) 343-4342
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2238 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-023-324-3


  Query Match             82.0%;  Score 16.4;  DB 4;  Length 2238;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcc 18
           |||||| ||||||||||
Db      34 CCGCGCCCGGCCGCAGCC 17


RESULT  15
US-08-910-820-7/c
; Sequence 7, Application US/08910820
; Patent No. 6258579
;  GENERAL INFORMATION:
;     APPLICANT:  Mercurio, Frank
;     APPLICANT:  Zhu, Hengyi
;     APPLICANT:  Barbosa, Miguel
;     APPLICANT:  Li, Gian
;     APPLICANT:  Murray, Brion W.
;     TITLE OF INVENTION:  STIMULUS-INDUCIBLE PROTEIN KINASE
;     TITLE OF INVENTION:  COMPLEX AND METHODS OF USE THEREFOR
;     NUMBER OF SEQUENCES:  25
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SEED and BERRY LLP
;       STREET:  6300 Columbia Center, 701 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/910,820
;       FILING DATE:  12-AUG-1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Maki, David J.
;       REGISTRATION NUMBER:  31,392
;       REFERENCE/DOCKET NUMBER:  860098.413C1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2251 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-910-820-7


  Query Match             82.0%;  Score 16.4;  DB 4;  Length 2251;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ccgcgctcggccgcagcc 18
           |||||| ||||||||||
Db      47 CCGCGCCCGGCCGCAGCC 30


Search completed: June 23, 2002, 15:03:00
Job time: 78204 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 06:31:23 ; Search time 2161.72 Seconds
(without alignments)
232.332 Million cell updates/sec

Title: US-09-747-514A-2
Perfect score: 24
Sequence: 1 ggtagcgcccgtaacgatcggccg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	24	100.0	3123	1	PAU38241	U38241 Pseudomonas
c 2	24	100.0	13075	1	AE004946	AE004946 Pseudomon
c 3	19.2	80.0	110000	2	LMFLCHR16_03	Continuation (4 of
c 4	18.8	78.3	346897	1	AP002995	AP002995 Mesorhizo
c 5	18.4	76.7	1993	1	AF417513	AF417513 Sinorhizo
c 6	18.4	76.7	299350	1	SME591786	AL591786 Sinorhizo
c 7	17.6	73.3	927	6	AX123252	AX123252 Sequence
c 8	17.6	73.3	16985	1	AE004594	AE004594 Pseudomon
9	17.6	73.3	43903	2	AC004396	AC004396 Pseudomon
10	17.6	73.3	309400	6	AX127153	AX127153 Sequence
c 11	17.2	71.7	11208	1	AE009268	AE009268 Agrobacte
12	17.2	71.7	12819	1	AE008344	AE008344 Agrobacte
13	16.8	70.0	11802	1	AE000030	AE000030 Mycoplasm
14	16.8	70.0	12850	1	AE008205	AE008205 Agrobacte
c 15	16.8	70.0	12952	1	AE009408	AE009408 Agrobacte
c 16	16.8	70.0	22838	1	AF081920	AF081920 Pseudomon
17	16.8	70.0	51233	2	AC110145	AC110145 Rattus no
c 18	16.6	69.2	136	9	HSU64481	U64481 Human rheum
c 19	16.6	69.2	560	6	AX310886	AX310886 Sequence
c 20	16.6	69.2	3547	1	AF247710	AF247710 Sinorhizo
21	16.6	69.2	4207	1	AF290948	AF290948 Pseudomon
c 22	16.6	69.2	8557	1	AE008212	AE008212 Agrobacte
c 23	16.6	69.2	10009	1	AE005628	AE005628 Escherich
24	16.6	69.2	10981	1	AE005890	AE005890 Caulobact
25	16.6	69.2	11742	1	AE009402	AE009402 Agrobacte
26	16.6	69.2	12772	1	AE004727	AE004727 Pseudomon
c 27	16.6	69.2	21327	1	SCL11	AL157953 Streptomy
c 28	16.6	69.2	35961	1	MLU15181	U15181 Mycobacteri
c 29	16.6	69.2	38503	1	MSGB1912CS	L01536 M. leprae g
c 30	16.6	69.2	42001	1	AF311901	AF311901 Citrobact
31	16.6	69.2	44872	5	AC091773	AC091773 Takifugu
c 32	16.6	69.2	300000	1	SME591784	AL591784 Sinorhizo
c 33	16.6	69.2	318703	1	AP002567	AP002567 Escherich
c 34	16.6	69.2	332635	1	AP003005	AP003005 Mesorhizo
35	16.6	69.2	346820	1	AP003008	AP003008 Mesorhizo
c 36	16.6	69.2	348450	1	MLEPRTN4	AL583920 Mycobacte
c 37	16.2	67.5	1386	6	AX203076	AX203076 Sequence
c 38	16.2	67.5	2157	9	AB056761	AB056761 Macaca fa
39	16.2	67.5	5660	6	AX347259	AX347259 Sequence
c 40	16.2	67.5	10449	1	AE005449	AE005449 Escherich
c 41	16.2	67.5	14029	1	AE007013	AE007013 Mycobacte
42	16.2	67.5	15898	1	AX024393	AX024393 Sequence
43	16.2	67.5	15898	6	AX024286	AX024286 Sequence
44	16.2	67.5	22775	1	SC5G8	AL353872 Streptomy
c 45	16.2	67.5	33522	1	MTCY02B10	Z75555 Mycobacteri

ALIGNMENTS

RESULT 1
PAU38241/c
LOCUS PAU38241 3123 bp DNA linear BCT 05-OCT-1996
DEFINITION Pseudomonas aeruginosa orotate phophoribosyl transferase (pyrE), catabolite repression control protein (crc) and RNasePH (rph) genes, complete cds.
ACCESSION U38241 L12038
VERSION U38241.1 GI:1079660
KEYWORDS .
SOURCE Pseudomonas aeruginosa strain=PAO1.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 3123)
AUTHORS MacGregor,C.H., Arora,S.K., Hager,P.W., Dail,M.B. and Phibbs,P.V. Jr.
TITLE The nucleotide sequence of the Pseudomonas aeruginosa pyrE-crc-rph region and the purification of the crc gene product
JOURNAL J. Bacteriol. 178 (19), 5627-5635 (1996)
MEDLINE 96421989
REFERENCE 2 (bases 1 to 3123)
AUTHORS Hager,P.W. and Phibbs,P.V. Jr.
TITLE Direct Submission

```
  JOURNAL   Submitted (11-OCT-1995) Paul W. Hager, Microbiology & Immunology,
            East Carolina University, Greenville, NC 27858, USA
COMMENT     On Nov 29, 1995 this sequence version replaced gi:496210.
FEATURES             Location/Qualifiers
     source          1. .3123
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
                     /map="11 min"
     gene            complement(414. .1055)
                     /gene="pyrE"
     CDS             complement(414. .1055)
                     /gene="pyrE"
                     /EC_number="2.4.2.10"
                     /function="catalytic activity: orotidine-5'-phosphate +
                     pyrophosphate = orotate + 5-phospho-alpha-d-ribose
                     1-diphosphate; Method: conceptual translation supplied by
                     author."
                     /codon_start=1
                     /transl_table=11
                     /product="orotate phosphoribosyl transferase"
                     /protein_id="AAC44427.1"
                     /db_xref="GI:1079661"
                     /translation="MQAYQRDFIRFAIERGVLRFGEFTLKSGRTSPYFFNAGLFDSGL
                     ALARLGRFYAEAVIDSGIDFDVLFGPAYKGIPLAATTAVALAEQHQRDLPWCFNRKEA
                     KEHGEGGTLVGAPLSGRVLIIDDVITAGTAIREVMQIIDAQGARAAGVLIALNRQERG
                     KGELSAIQEVERDFGMPVVSIVSLEQVLEYLAEDAELKKHLPAVEAYRAQYGI"
     gene            1136. .1915
                     /gene="crc"
     CDS             1136. .1915
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAC44428.1"
                     /db_xref="GI:1079662"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(2394. .3113)
                     /gene="rph"
     CDS             complement(2394. .3113)
                     /gene="rph"
                     /note="similar to ribonuclease PH"
                     /codon_start=1
                     /transl_table=11
                     /product="RNase PH"
                     /protein_id="AAC44429.1"
                     /db_xref="GI:1079663"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
                     ESGVPRFLKGQGQGWLTAEYGMLRRSTGERNQREASRGKQGGRTLEIQRLIDRSLRAA
                     LDLSKLGENTLYIDCDVIQTDGGTRTASITGATVSLIDALAVLKKRVALKGNPLKQIV
                     AAVSVGIYQGVPVLDLDYLEDSAAESDLNVVMTDCRRFIEVQGTAEGAPFRPAELNAM
                     LELAQQGMQELVRTQRAALAE"
BASE COUNT      551 a   1020 c    989 g    563 t
ORIGIN


  Query Match           100.0%;  Score 24;  DB 1;  Length 3123;
  Best Local Similarity  100.0%;  Pred. No. 1.4;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggtagcgcccgtaacgatcggccg 24
          ||||||||||||||||||||||||
Db   1418 GGTAGCGCCCGTAACGATCGGCCG 1395


RESULT   2
AE004946/c
LOCUS       AE004946               13075 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 507 of 529 of the complete
            genome.
ACCESSION   AE004946 AE004091
VERSION     AE004946.1  GI:9951650
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .13075
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            71. .850
                     /gene="crc"
                     /note="PA5332"
     CDS             71. .850
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAG08717.1"
                     /db_xref="GI:9951651"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(909. .1280)
                     /gene="PA5333"
     CDS             complement(909. .1280)
                     /gene="PA5333"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08718.1"
                     /db_xref="GI:9951652"
                     /translation="MNDVSEEVAATPGQEARQWAMFCHFAAFIGLVFPFGSLLGPLIV
                     WQLKRESDPFVDAQGKEALNFQITVALAMFVSFLLMLVVIGFFLLGLVCLAALVLTII
                     AGIKANEGKAYRYPFCLRLIK"
     gene            complement(1327. .2046)
                     /gene="rph"
                     /note="PA5334"
     CDS             complement(1327. .2046)
                     /gene="rph"
                     /codon_start=1
                     /transl_table=11
                     /product="ribonuclease PH"
                     /protein_id="AAG08719.1"
                     /db_xref="GI:9951653"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
```

comp

```
                ESGVPRFLKGQGQGWLTAEYGMLPRSTGERNQREASRGKQGGRTLEIQRLIGRSLRAA
                LDLSKLGENTLYIDCDVIQADGGTRTASITGATVALIDALAVLKKRGALKGNPLKQMV
                AAVSVGIYQGVPVLDLDYLEDSAAETDLNVVMTDAGGFIEVQGTAEGAPFRPAELNAM
                LELAQQGMQELFELQRAALAE"
gene            2227. .3090
                /gene="PA5335"
CDS             2227. .3090
                /gene="PA5335"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG08720.1"
                /db_xref="GI:9951654"
                /translation="MVHSMTAFARVEQAGTHGTLSWELRSVNHRYLEPHLRLPDAFRD
                LEGAVREALRQGLSRGKVECTLRFAEETAGKSLQVDQERARQLVAAAEGVAALIRQPA
                PLDPLAVLAWPGVLVADSADPQALNAAALEAFGQALEQLKAGRSREGLELAKLLNDRL
                DAMLVEVGNLRELVPTMLANQRQKILDRFAELKAELDPQRLEQELVLLAQKSDVAEEL
                DRLATHVGEVRRVLKAGGAAGRRLDFLMQELNREANTLGSKAFDPRSTQAAVNLKVLI
                EQMREQVQNIE"
gene            3149. .3760
                /gene="gmk"
                /note="PA5336"
CDS             3149. .3760
                /gene="gmk"
                /codon_start=1
                /transl_table=11
                /product="guanylate kinase"
                /protein_id="AAG08721.1"
                /db_xref="GI:9951655"
                /translation="MSGTLYIVSAPSGAGKTSLVKALLDAAPEVRVSVSHTTRGMRPG
                EVDGVNYHFTSREEFLAMLERNEFLEHAEVFGNLYGTSQRWVEKTLAEGLDLILEIDW
                QGAQQVRRLMPEAQSIFILPPSQEALRQRLTNRGQDSDEVIERRMREAVSEMSHYVEY
                DHLVINDDFAHALDDLKAIFRARQLRQDAQQQRHAELLGRLLA"
gene            3849. .4115
                /gene="rpoZ"
                /note="PA5337"
CDS             3849. .4115
                /gene="rpoZ"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase omega subunit"
                /protein_id="AAG08722.1"
                /db_xref="GI:9951656"
                /translation="MARVTVEDCLDNVDNRFELVMLATKRARQLATGGKEPKVAWEND
                KPTVVALREIASGLVDENVVQQEDIVEDEPLFAAFDDEANTEAL"
gene            4181. .6286
                /gene="spoT"
                /note="PA5338"
CDS             4181. .6286
                /gene="spoT"
                /codon_start=1
                /transl_table=11
                /product="guanosine-3',5'-bis(diphosphate)
                3'-pyrophosphohydrolase"
                /protein_id="AAG08723.1"
                /db_xref="GI:9951657"
                /translation="MPGIDAFADRLSSYLGPDQVNLVRRAYYYAEQAHDGQRRRSGEP
                YVTHPLAVSNILADMHMDHQSLMAAMLHDVIEDTGIAKEALTAQFGETVSELVDGVSK
                LTQMNFETKAEAQAENFQKMAMAMARDIRVILVKLADRLHNMRTLEVLSGEKRRRIAK
                ETLEIYAPIANRLGMHTMRVEFEDLGFKAMHPMRAERIRQAVKKARGNRREIVGKIQE
                SLVNCLAREGMEGQVIGREKHLYSIYQKMRGKRKAFNEIMDVYAFRIIVDKVDTCYRV
                LGAVHSLYKPFPGRFKDYIAIPKANGYQSLHTTLFGMHGVPIEIQIRTREMEEMANHG
                IAAHWLYKSNDETPKGTHARARQWVKGVLELQQRAGNSLEFIENVKIDLFPDEVYVFT
                PKGRIMELPKGSTAVDFAYAVHTDVGNSCIACRINRRLAPLSEPLQSGQTVEIVTAPG
                ARPNPAWLSFVVTGKARTHIRHALKLQRRSESINLGERLLNKTLTGFDSHLEKIPQER
                IQAVLAEYRMEVFEDLLEEIGLGNRMAYVVARRLLSSEGEEAPSAEGPLAIRGTEGLV
                LSYAKCCTPIPGDPIVGHLSAGKGMVVHLESCKNISEVRHNPEKCIQLSWAKDVTGEF
                NVELRVELEHQRGLIALLATSVNAADGNIEKISMDERDGRISVVQLVVSVHDRVHLAR
                VIKKLRALKGVIRITRLRS"
gene            6348. .6728
                /gene="PA5339"
CDS             6348. .6728
                /gene="PA5339"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG08724.1"
                /db_xref="GI:9951658"
                /translation="MTKTVIHTDNAPAAIGTYSQAIKAGNTVYVSGQIPLDPKTMELV
                EGFEEQTVQVFENLKAVIEASGGYLHDVAKLNIFLTDLSHFAKVNEIMGRYFTQPYPA
                RAAIGVAALPKGAQVEMDAIVVLE"
gene            6783. .7514
                /gene="PA5340"
CDS             6783. .7514
                /gene="PA5340"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG08725.1"
                /db_xref="GI:9951659"
                /translation="MRRVIFLAAAATLLAGCAGTADPSGTWINQAAIDAASKDGKLRE
                ALLAYGPNLEWKLDSKAGEATFSNGFELGEGTLSKSDDEHWKVAFYGDDNQESLELDG
                KELIQQASANGPEQRFRRLDPQPAASSPAGSGFERALYGSYLKGSWKIREGQGQGGKV
                EFQANGLVSGLPGAERYALCLAGDCAAMSGDNDSIWLQQGNRGRELLFSLDDDELQLF
                EAVNTAGANEMPSYVPGKRVWLLER"
gene            complement(7521. .8141)
                /gene="PA5341"
CDS             complement(7521. .8141)
                /gene="PA5341"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG08726.1"
                /db_xref="GI:9951660"
                /translation="MFAAFVLVASTHFAALLSPGPDFFLLVRAALLRGRRQADGVAAG
                IALANLLSMLLVLGLLASVPDSAHSALRWLQALGGLYFLWLAGQALLAQRRLEMPAQR
                DVPSRGGYLRGLRDGLLASSLNPKLPIFYAGLFGVLARFSLPGWALALCLAWMSLAVL
                CWDMALVRLLDRPRWRGWLQRRVGALDRLCGVLLLALGGWLVLAAL"
gene            complement(8208. .9008)
                /gene="PA5342"
CDS             complement(8208. .9008)
                /gene="PA5342"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG08727.1"
                /db_xref="GI:9951661"
                /translation="MAANRLLARPWLPGVELFHADFSGQAFGRHSHSAFAIGSILHGV
                GGYQCRGRRHALPAGTLSLMNPEEPHTGHAESERLVYRMLYVEETRLLGLLGRKQLPR
                GFGELNPADDGSVAAGLAWLAEDFQRSDALGLESRLLALLEQVFVRHGGLRPAAAAPR
                DGGTTARLRDYLEAHATEALGLDQLAQLLQRHPRHLIEAFRRAYGVPPHTYQLQCRVR
                QAKRLLLEERPLAELAQHLGFYDQAHFSTTFRRFTGVTPGQFRRSAKT"
gene            complement(9008. .9859)
                /gene="PA5343"
CDS             complement(9008. .9859)
```

Query Match 100.0%; Score 24; DB 1; Length 13075;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Comp

```
Qy      1 ggtagcgcccgtaacgatcggccg 24
          ||||||||||||||||||||||||
Db    353 GGTAGCGCCCGTAACGATCGGCCG 330
```

RESULT 3
LMFLCHR16_03/c
WPCOMMENT
Sequence split into 11 fragments LOCUS LMFLCHR16 Accession AL499619

Fragment Name	Begin	End
LMFLCHR16_00	1	110000
LMFLCHR16_01	100001	210000
LMFLCHR16_02	200001	310000
LMFLCHR16_03	300001	410000
LMFLCHR16_04	400001	510000
LMFLCHR16_05	500001	610000

```
     LMFLCHR16_06       600001    710000
     LMFLCHR16_07       700001    810000
     LMFLCHR16_08       800001    910000
     LMFLCHR16_09       900001   1010000
     LMFLCHR16_10      1000001   1030105
Continuation (4 of 11) of LMFLCHR16 from base 300001 (AL499619 Leishmania major chromosc


  Query Match             80.0%;  Score 19.2;  DB 2;  Length 110000;
  Best Local Similarity   87.5%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           |||||||||||||  || ||||| |
Db   16300 GGTAGCGCCCGTAGCGGTCGGCTG 16277


RESULT   4
AP002995/c
LOCUS       AP002995               346897 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 2/21.
ACCESSION   AP002995 BA000012
VERSION     AP002995.2  GI:14021442
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 346897)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994963.
FEATURES             Location/Qualifiers
     source          1. .346897
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     tRNA            74. .160
                     /product="tRNA-OTHER"
     gene            complement(240. .1109)
                     /gene="msl8587"
     CDS             complement(240. .1109)
                     /gene="msl8587"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB48055.1"
                     /db_xref="GI:14021443"
                     /translation="MYDQTQTTGITALPYTVVRIRAGDVANPQAVHKGTGFFYHFREG
                     AGNVPLIVSNKHVLVGKEWLEFDFALLDNQGKRVFGPANTFRLTKNQLPIFEHPDPTV
                     DLAAIPIQPIVEAMRGQGKEIFNLLLGRETIAPDNIQAILHAATSVLMVGFPSGIMDE
                     KNNLPVVRRGILAVNYLTDYRGETNFVVDIAAFGGSSGSPVFAFFENMLPDGSGGVTM
                     LAQPQVYFLGVLHSGPSMTATGKIVPVPVPTSDQIAQTQMMLHLGYCVKASRVEELVE
                     TIRAKISEPTASG"
     gene            complement(1118. .1657)
                     /gene="mll0476"
     CDS             complement(1118. .1657)
                     /gene="mll0476"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB48056.1"
                     /db_xref="GI:14021444"
                     /translation="MFPISALIGFEKDGQWRGALALAGEAMAQTAEDNFRIEVWDREE
                     KNLVETICRSPDSFISQAAWQAAIRRRPGMVLIHYNSRHVMEKIITPGEPKVPPQTIV
                     DGSVHAGQDVALGDLREWHKLRAWCKNCSHHAEVKPAALIKRYGEAALFSTVERALFC
                     TGCDRGGPVRLEIHKLPRN"
     gene            2193. .2558
                     /gene="mlr0478"
     CDS             2193. .2558
                     /gene="mlr0478"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB48057.1"
                     /db_xref="GI:14021445"
                     /translation="MALALGGALMKWRLYFVVMTLAVTSGCAADYLNNYDTVTLASGD
                     ANHTNSLLQTVDPFNPNSNNTHIEGDGQRIVGVVQRYRSPISAGGGYSGNCPTDNSVA
                     ADGSHCGKRSAESRPGGAM"
     gene            2651. .3412
                     /gene="mlr0479"
     CDS             2651. .3412
                     /gene="mlr0479"
                     /codon_start=1
                     /transl_table=11
                     /product="succinoglycan biosynthesis protein; ExoI"
                     /protein_id="BAB48058.1"
                     /db_xref="GI:14021446"
                     /translation="MPLRLVLVTVAAASALATQFVMHNGGSLPEINLQNLIKPKPVAI
                     SGVASVIDGDTIEMHGQRVRFNGIDAPESAQYCDDAKGFEYPCGRRSAEALDGFLAAS
                     KPVQCAFVTWDRYGRFVGDCTRADGSSVAAWMVEHGQALDWPKYSNGAYAAQQAKAKA
                     AKVGLWIGSFQAPWDWRAQHSDDAQTPTAPLFATGNGNPGCNIKGNISASGERIYHLP
                     GQKYYAVTVISQNKGERWFCSEAEAVAAGWRRSKR"
     gene            3429. .3782
                     /gene="mlr0480"
     CDS             3429. .3782
                     /gene="mlr0480"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB48059.1"
                     /db_xref="GI:14021447"
                     /translation="MKPHVRAAVAAIALSHSSGRNVSSVYSYSELKYVNIDASVQGNV
                     VDAYDYNNSCHIDGTLPSLYHYGQNSHIDLQPQSGGQYDGYDYGSNSHFELTVSGNSA
                     ELYDYGEGGFFSFST"
     gene            complement(5516. .5947)
                     /gene="mll0481"
     CDS             complement(5516. .5947)
                     /gene="mll0481"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB48060.1"
                     /db_xref="GI:14021448"
                     /translation="MDNSDRTGKDPSTGRFLPGHSGFKKVGSRSRVSRQAIDTANEHA
                     ATAMTMLVGKMLAGDIRAAEILLRYGADPFLPKNRTVELPDVENPTSENLATAVLSAA
                     WSGEISPDEAKQLLDAIAAGKGIIEDAELLRRVEQLEAVKR"
     gene            complement(6065. .6883)
                     /gene="mll0482"
     CDS             complement(6065. .6883)
                     /gene="mll0482"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB48061.1"
                     /db_xref="GI:14021449"
                     /translation="MIKDFKRIDGASSRGSLFQWNYNGSVISAFWPNEYSTMSPDIDA
                     AASALNNGVLAAGNAVIGAAERLKRTDEIRDVSLEAIARYVAAPFRYLRDRAIDERNS
                     VGISRNSLAQPIQIPNPTPFSIAADSLERQEVRRMLDGFDQIGEAFDWVMQSPGITVL
                     NAVAPSIDLTIFSKDQQLAGLIVEEFAIRKWARTIDNPSNDNLATAANPLKNKMSDDL
```

```
                VRDLAKARLDGLKLREQAVSDAEKLLLAVIAFASQAGNLTPDEAFKMLMGQKVA"
gene            complement(7133. .7753)
                /gene="mll0483"
CDS             complement(7133. .7753)
                /gene="mll0483"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48062.1"
                /db_xref="GI:14021450"
                /translation="MANYVILRTLVMKNRASTITADHANFAKLGQIADHAFLWLSGRT
                REWASPVNPSEFQYRDECLDMVSGYEGNIRTTFSFDKTYQSYSFMFDAATTDGHPFVV
                WGTLGGCADNMVVSSTHFSFGIINGVMSASVDQSDNWQFEDSSSTVVSEMLDRHLQPW
                RDFPTGVARQSKELPPAIIRALVKLGYRSKDTRRWTAPNQFAENIH"
gene            complement(7827. .10169)
                /gene="mll0485"
CDS             complement(7827. .10169)
                /gene="mll0485"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48063.1"
                /db_xref="GI:14021451"
                /translation="MRDGSSVRLPPELLPYRQFVVWRLEQKPGSPKPAKMPYSANGGR
                ADTTNPATWSDYPTAVRALASGQYAGIGFVFAENDPFFVIDLDGCRDATTGELSAEAA
                AITRMFPNAAWELSQSGEGIHIFGRCDKVALGGFRNKWAGWCEFYRTKRFVAFGPGGI
                QGNIDIDWTNQLAILVPARGVGEQDASQEAVERDPRWIGPDDDEELLRRFLASTGGTK
                AKLQASPTNLQLWEAHAAALAQYFPSATGKPFDHSSADQALANVLAFWTGRDYGRIIR
                LFGRSALGRREKWTRRAAYRNSTVNAAVRSCQTVYGGNGQERRSRQVQENVRIGSDIG
                ADIHTRVMTLEDMLQDLWFIRHSKSGGAIADSRTLNVLPFATAQREYSASRYEVAAST
                IDPKTGDFKKRYLPVFNQWEQSPSRKSVDVITWKPNGAVICEAPESQGAGFNTWRGFQ
                PYRLPDGSQSLVEVWSEHLAYLIPIEAERVRFEQWLAHIVQRPEELPQSAYLFITEET
                GIGRNWLSSVLVRVLRGYVAAGIDIQEMLDGTFNGRLSRKLLAVVDEAKAGMQGYGRW
                SHSEKLKSMINPEMRRINEKHGLEFVEWNCCRWLFFSNNWDALPFDASDRRFNVVANP
                THRQSPDYYERIYSQIANPDFISAVRKRLELLDISGFKPGAIPDMNEAKRKALGAMAS
                DLDNLLAAFRDEWPGPLAERSDVINYLHVSIAKLPGEKTINKLIEQAGMKLVERISIG
                PNKYRFVVVRDYTAQQIETNPQWAYAEAVNARGKYGFRQG"
gene            complement(10162. .10392)
                /gene="mll0486"
CDS             complement(10162. .10392)
                /gene="mll0486"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48064.1"
                /db_xref="GI:14021452"
                /translation="MECVAMAGNLRVREMNQSAIVLSIPQVCATLGISRPTLDRYIKD
                NPSFPRKKKIGPRLVGFLRSDIETYIGSLPDA"
gene            complement(10595. .11803)
                /gene="mll0487"
CDS             complement(10595. .11803)
                /gene="mll0487"
                /codon_start=1
                /transl_table=11
                /product="probable prophage integrase"
                /protein_id="BAB48065.1"
                /db_xref="GI:14021453"
                /translation="MLTIREIQSLVHKGEPGQWSHTKALYFFIAPSGRPFWAIRYMLK
                GRRRQMSLTDSIPATVRELKQIELLSETEREKVRKGIDPLAEREAAKAASPEPVRQLV
                TFKSVALDYIAANRGGWKNAKHVQQWENTLATYVYPVIGHLEPQSITTAEVLKILQPL
                WATKRETASRVRSRVETIIGAAKAKGMSEKPEIWANHHNPARWADNLEFWLNGKQEKN
                HFSAMPYPEIPAFAQSIAEREDYSTKALQFLILTATRTNETLGAEWSEIDTEKAVWTI
                PASRMKSGKEHRVPLSSEALKLVASLLRIEGNPYLFPGAIRNKSLSNMALLMVMRGLK
                LGHYVPHGFRSAFRDWAAEETVYPNHIVEMALAHTIKNKVEAAYRRGDLMEKRRQLME
                DWSAYLASAR"
gene            12686. .13279
                /gene="mlr0488"
CDS             12686. .13279
                /gene="mlr0488"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48066.1"
```

```
  Query Match           78.3%;  Score 18.8;  DB 1;  Length 346897;
  Best Local Similarity  90.9%;  Pred. No. 2.6e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ggtagcgcccgtaacgatcggc 22
          ||| ||||||||| ||||||||
Db 318216 GGTGGCGCCCGTATCGATCGGC 318195


RESULT   5
AF417513/c
LOCUS       AF417513               1993 bp    DNA     linear   BCT 30-JAN-2002
DEFINITION  Sinorhizobium meliloti inorganic pyrophosphatase gene, partial cds.
ACCESSION   AF417513
VERSION     AF417513.1  GI:18419443
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 1993)
  AUTHORS   Jumas-Bilak,E., Michaux-Charachon,S. and Teyssier,C.
  TITLE     High prevalence of the H+ proton-pumping inorganic pyrophosphatase
            gene in alpha proteobacteria and evidence of lateral transfer in
            its phylogeny
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1993)
  AUTHORS   Jumas-Bilak,E., Michaux-Charachon,S. and Teyssier,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) INSERM 431, Faculte de Medecine, Avenue
            Kennedy, Nimes 30900, France
FEATURES             Location/Qualifiers
     source          1. .1993
                     /organism="Sinorhizobium meliloti"
                     /strain="2011"
                     /db_xref="taxon:382"
     CDS             <1. .>1993
                     /codon_start=1
                     /transl_table=11
                     /product="inorganic pyrophosphatase"
                     /protein_id="AAL69329.1"
                     /db_xref="GI:18419444"
                     /translation="MQEVAGFIREGAQAYLTRQYTTIAIVGVVVFIAAWLLLSGAAAI
                     GFLIGAVLSGAAGFIGMHVSVRANVRTAQAASVSLASDIDIAFKSGAITGMLVAGLAL
                     LGVSVYIYFVLIVWLGYAPADRAVIDSLVALGFGASLIAIFVRLGGGIFTKGADVGGDL
                     VGKVEACIREDDPCIPATIANNVGHNVCGCAAMAADLFETYAVTVVATMVLASIFFAG
                     APVLATVMTYPLAICAACIITSIIGTFFVKLGANASIMGALYRGLIVTGALSILGLGA
                     ATSLTIGWGSIGTVAGMDITGWNLFLCGIIGLIVTALIVVITEYYTGTNKRPVNSIAQ
                     ASVTGHGTNVIQGLAVSLESTALPAIVIVGGIIATYQFAGLFGTAIAVTAMLGLAGMI
                     VALDAFGPVTDNAGGIAEMSHLPPEVRKSNDALDAVGNTTKAVTKGYAIGSAGLGALV
                     LFAAYSNDLAYFAANGDKHPYFADVGTISFDLSNPYVVAGLIFGGLIPYLFGGIAMTA
                     VGRAGSAVVEEVRRQFKEKPGIMEGKDRPDYGRAVDMLTKAAIREMIIPSLLPVLAPI
                     VVYFGVLLISGSKASAFAALGASLLGVIVNGLFVAISMTSGGGAWDNAKKSFEDGFVD
                     RNGTRHMKGSEAHKASVTGDTVGDPYKDTAGPAVNPAIKI"
BASE COUNT      317 a    656 c    606 g    414 t
ORIGIN


  Query Match           76.7%;  Score 18.4;  DB 1;  Length 1993;
  Best Local Similarity  95.0%;  Pred. No. 5.2e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 agcgcccgtaacgatcggcc 23
          |||||||||||||||| |||
Db    804 AGCGCCCGTAACGATCAGCC 785


RESULT   6
SME591786/c
LOCUS       SME591786              299350 bp    DNA     linear   BCT 16-AUG-2001
DEFINITION  Sinorhizobium meliloti 1021 complete chromosome; segment 5/12.
ACCESSION   AL591786 AL591688
```

```
VERSION     AL591786.1  GI:15073988
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 299350)
  AUTHORS   Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
            Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
            Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
            Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
            Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
  TITLE     From the Cover: Analysis of the chromosome sequence of the legume
            symbiont Sinorhizobium meliloti strain 1021
  JOURNAL   Proceedings of the National Academy of Sciences of the United
            States of America. 98 (17), 9877-9882 (2001)
   PUBMED   11481430
REFERENCE   2  (bases 1 to 299350)
  AUTHORS   Gouzy,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
            EU Consortium
COMMENT     MELILO EU Consortium:
            Laboratoire de Biologie Moleculaire des Relations
            Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
            France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
            Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
            France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
            Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
            D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
            Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
            B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
            des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
            B-5030 Gembloux, Belgium.  E-mail:Jerome.Gouzy@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/meliloti.html.
FEATURES             Location/Qualifiers
     source          1. .299350
                     /organism="Sinorhizobium meliloti"
                     /strain="1021"
                     /db_xref="taxon:382"
     gene            complement(132. .608)
                     /gene="SMc02401"
     CDS             complement(132. .608)
                     /gene="SMc02401"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAC45636.1"
                     /db_xref="GI:15073989"
                     /translation="MKTLAIVALSLATVISSVPPAEAFPIVPAPKPQAADVQLAQFPD
                     RARCPGVGCPRMGRERWRGNRHNYRGNRYYGNRYHHRRYYRDDDDDLGAFFGGLAAGA
                     IVGGVLSQPRYAAPRYSGGGNAHVEWCYSRYRSYRAWDNTFQPYNGPRRQCYSPYS"
     gene            complement(788. .2833)
                     /gene="SMc02403"
     CDS             complement(788. .2833)
                     /gene="SMc02403"
                     /EC_number="3.2.1.-"
                     /function="structural elements; cell envelope; murein
                     sacculus, peptidoglycan"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE MUREIN TRANSGLYCOSYLASE PROTEIN"
                     /protein_id="CAC45637.1"
                     /db_xref="GI:15073990"
                     /translation="MMRGTVFLPVAAMIGAGMLAFSADAAEGRIPVPTSKPDARKIGS
                     LETGELSTILGADLKSGLDALSDRQPEKAIAVRDRMPAGTLDRHILAWAIAVSGQKGI
                     PSYDIAAAQRELQGWPGLKSLRSHSERALYRENAPAADVIAAFGTTRPETAEGAIILA
                     RALAAQGQSSAAATHLRAFWLSEALDKDTETKILSEFSDLLTTEDHRQRMEMLLYRTR
                     IDQAERFSDLGKAQSLFRAWAAVTRGSSNAASLIEAVDASWRSKPSYLFIRIEHLRKQ
                     QKYAEAARLLDQMPKDDGALVNPGEWWVEQRIVSRGLLDAGDFRGAYRIAANHAAKRA
                     TDLVDAEFHAGWYALRGLEEPETAALHFRRILEASSRPISASRAWYWLGRAAEAGGPG
                     DAREYFTNAARYPATFYGQLAAARLGQAKLDMSYPSPTQEDRIRLEGREAVKAIGRLE
                     AAGHGWRAESLYRALAEELTSPGELAILAERAEKSGNHQLSLQIGKIAFGRGIEVAAL
                     AFPIGVIPPSADIDGAGKALAYAIARQESAFNPAAISVANARGLLQLLPGTAKGVASR
                     YGLAYSKDRLTSDAGYNATLGAHYLGEQIDSFGGSYILTFIAYNAGPRRVPDWLARYG
                     DPRGKPIDEVVDWIERIPFEETRNYVQRVMENYQVYKSRLGQDTDIVADLRLGRALP"
     gene            3130. .4014
                     /gene="dapA1 OR SMc02404"
     CDS             3130. .4014
                     /gene="dapA1 OR SMc02404"
                     /EC_number="4.2.1.52"
                     /function="small molecule metabolism; amino acid
                     biosynthesis; lysine"
                     /note="Product confidence : putative
                     Gene name confidence : putative
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE DIHYDRODIPICOLINATE SYNTHASE PROTEIN"
                     /protein_id="CAC45638.1"
                     /db_xref="GI:15073991"
                     /translation="MFKGSIPALVTPFTAAGSVDADSFVAHVEWQIKEGSHGLVPVGT
                     TGESPTLSHDEHKKVVELCVEASARRVPVIAGAGSNNTIEAIELAQHAEKAGADAILV
                     VTPYYNKPTQKGLFAHFAAIAESVKLPIVIYNIPGRSVVDMSVETMAALAKAHPTIVG
                     VKDATGRIERVSEQRMACGKAFVQLSGEDATALGFNAHGGVGCISVTANVAPRLCAEF
                     QEATLAGNYAKALELQDKLMPLHKAIFLEPGVCGAKYALNRLGRMSFTVRSPLLSALE
                     PATASAIDAALRHAGLMN"
     gene            4014. .4493
                     /gene="smpB OR SMc02405"
     CDS             4014. .4493
                     /gene="smpB OR SMc02405"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : putative
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="CONSERVED HYPOTHETICAL PROTEIN"
                     /protein_id="CAC45639.1"
                     /db_xref="GI:15073992"
                     /translation="MAPKGSERTVKKVVAENRKARFNYEIVDTYEAGLVLTGTEVKSL
                     REGKANIAESYATDEGGEIWLINSYLPEYLQANRFNHETRRRRKLLLSKREVNRLQGA
                     VNREGMSLIPLRIYFNERGRVKLELALGKGKKLHDKRETSKERDWNRQKNRLLKERG"
     repeat_region   complement(4651. .4820)
                     /note="Sm-4 OR SMc04541
                     REPEAT SM-4
                     predicted by Homology"
                     /evidence=not_experimental
     repeat_region   complement(4919. .5085)
                     /note="Sm-4 OR SMc04542
                     REPEAT SM-4
                     predicted by Homology"
                     /evidence=not_experimental
     gene            complement(5081. .9316)
                     /gene="SMc02406"
     CDS             complement(5081. .9316)
                     /gene="SMc02406"
```

```
                     /function="miscellaneous; hypothetical/partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC45640.1"
                     /db_xref="GI:15073993"
                     /translation="MIPAGPVLASVGFTLAIGAGAVSPAHAINECGPAIGGVVTCNTS
                     TYPKGPGIHIEYRIDTGLRATIDSVAVRWADDAGSDRPAIDIRSDKVTSEGDLIVTIE
                     GSTVTTGEPDRIGTRSDREVGEYADGLRVTQRGAGLVSATMLSGSIITYGYNAAGIYA
                     WQPGANENNKSDVIADFSGEDSTIETFGSYSQGIYGRTETGIGNVAVTAKDGTIVTHG
                     LSSIGVRGLKHSASLGDGDVTVAGDNITVTTGGNLSIGMEARVEDSKSGRGNASVFVR
                     NEAAIETSGTRAHGAQAFNLDPGDSTVVLEKSKITTRGEGAIGIRAEAGDGNAVTEVT
                     KGSIATAGASAIGLAAISQGNMGEARVSLDGGDVSTGGDNAEGLIATAKGAGGTATAR
                     LVDGNVKTAGRDAEGLIAHAGDESIPTSINADASVIMQAGSITTTGDGAGGMIAETDI
                     GPTPSTGKATAVQRAGTIMTSGGEIGGNEGSYAIAALSFGSGLASIEQAGGSATTAGA
                     QSHALYALSIFGNTMVTQAAGSSAVATGAKASGLRALAGPGGGNEVTLDGKVTAGSAP
                     AVHTIGSAGSKITIGSAAEIDGSASGTAIRDGDSDLDAADETGGNSTATLAGKVTGDV
                     VMGLGNDIINLTGGVLDGDVYGDDRLATAADGDDSFTWSGGTWLSSFFGGNGSDTATV
                     SAQSYDGTHHYLDGGDDVSTADGWTDRLNLSGVTASANGANIVNWEIVTLDAADVTIL
                     DGALEAGSEPETGLFLTNGSVLNGSDAFALAGNMAIDGTSQFIAFGEDIANGVYSISG
                     NLVNAGTVDMQDGGTGDVLSVGGNYTGAENSTVYIDTYLGDDASPTDHITVGGDTAGS
                     TAISIGNAGGPSAQTVAGIRVVGVAGASSGTFVLANANSEIKETGEAAITRGAYAYAL
                     RQVDNDDWFLQSTLAEDDTGPDTGGEQAVDPETEGGETVDPGTGGGETVDPGTGGGET
                     VDPGTGGGETVDPGTGGGETVDPGTGGGETVDPGTGGGETVDPGTGGGETVDPGTGGG
                     ETVDPETGGGETVDPGTGGGETVDPGTGGGETVDPGTGGGETVIPSYNQGVPVYETYA
                     RNLLGFIQLPTLRRRASGERVPALRSCGSVPLNVEGNQTAACAEDNTVWARIDGTRAH
                     VEPEESGARVSEYDSRSWLLQAGIDALLHSSGKGELIGGLTAHYGGINTDVSSFYGDG
                     SIDTRGYGLGATLTWYGQNGFYLDGQGQLTRFDSDLDSDILGRLADGNDGLGYAVSIE
                     AGHRIGLGDTWALTPQAQLVNTSVNFDDFVDPYGARVSLDDGNSLRGRIGLALDREQQ
                     WQGAEGDGRSSRLYGTIDLYREFLGDTRADVSGVSFGIEADDWIGEIGIGGSYNWGDD
                     KYSLYGEARASTGLSDFGGSYAVRASVGLKMKF"
     gene            5879. .6535
                     /gene="SMc02422"
     CDS             5879. .6535
                     /gene="SMc02422"
                     /function="miscellaneous; hypothetical/partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC45641.1"

  Query Match             76.7%;  Score 18.4;  DB 1;  Length 299350;
  Best Local Similarity   95.0%;  Pred. No. 4e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 agcgcccgtaacgatcggcc 23
           |||||||||||||||||| |||
Db  172938 AGCGCCCGTAACGATCAGCC 172919


RESULT   7
AX123252/c
LOCUS       AX123252                927 bp    DNA     linear   PAT 11-MAY-2001
DEFINITION  Sequence 3168 from Patent EP1108790.
ACCESSION   AX123252
VERSION     AX123252.1  GI:14040740
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 3168 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
BASE COUNT      187 a    253 c    277 g    210 t
ORIGIN


  Query Match             73.3%;  Score 17.6;  DB 6;  Length 927;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           ||| ||||||| ||| |||||| |
Db     687 GGTTGCGCCCGGAACAATCGGCTG 664


RESULT   8
AE004594/c
LOCUS       AE004594                16985 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 155 of 529 of the complete
            genome.
ACCESSION   AE004594 AE004091
VERSION     AE004594.1  GI:9947630
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 16985)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 16985)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .16985
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            303. .1883
                     /gene="PA1660"
     CDS             303. .1883
                     /gene="PA1660"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG05049.1"
                     /db_xref="GI:9947631"
                     /translation="MMHLLWPNYMRPLPAFSILQFDPLTRPGGALRVERDTPVESRPV
                     GEVACRFRTCYATEVLPLVLEALNYSVTGEGALLGLRLAMNCDGHLGELGLSRLRLHL
```

```
                AGERYISQMLYLCLLRNLESIQLIPLDAHGRPLDGVAGKPMSFRLTGERVQPVGFAED
                EALIPYPLNTFRGYRYLQEYFAFQDKFLFVDVDGLDLLGALPEETLKQVRGLELRFDI
                RKSGIQRLHPTLENVKLYCTPVANLFRHDALPIRLDGKQDEYLLMPAEYSLEDCGVFS
                VEGVTGWRPGGLGYQAYVPFESFEHDPSFDVPEARPHYSVRQRTSLLHDGLDTYLSFG
                IRHPEQIETLSVELTCTNQNLPQRLRLGDICVACEGTPEFLGFRNITPATPSYAPPLN
                RDFLWKLISNMSLNYLSLANVDALKVILETYDLPRYYDQHAAKVSKRLLGGLKAVRHE
                HVDRLHRGLPLRGLRTELTIDPEGYIGEGDLFVFASVLNEFFALYASINSYHELRVRS
                TQGDVYQWKPRMGLQPLL"
gene            1847. .2854
                /gene="PA1661"
CDS             1847. .2854
                /gene="PA1661"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05050.1"
                /db_xref="GI:9947632"
                /translation="METAHGFAAPAVSALARGIREYSLFQAVLLVMDRLRQEHPGLGD
                EALYDQLEFQANPSLGFPGSDVDRVEFFEERGMLRARLRFNLIGLFGASSPLPAFYSE
                QALGDSEEGNPTRNFLDLFHHRLHRLLLPIWRKYRYRVSFQSGASDAFSAQLFALVGL
                GGEEIRKARELNWKRLMPYLGLLSLRAHSAALIESVLRYYFKHAELVAEQCIERRVAI
                LEEQRNRLGLFNSQLGEDAVLGERVRDRSGKFRIHIRQLDWQRFHEFLPIGPGYQPLC
                ALVRFTLRDPLDYDIRLVLRHEEIREFRIGAQNACRLGWTSWLGRERADGVVTLGSKI
                H"
gene            2861. .5494
                /gene="PA1662"
CDS             2861. .5494
                /gene="PA1662"
                /codon_start=1
                /transl_table=11
                /product="probable ClpA/B-type protease"
                /protein_id="AAG05051.1"
                /db_xref="GI:9947633"
                /translation="MSMINVDLQQLIQALDADTRRDLEGSAERCVARGGSKVLVEDLL
                LGLLERPQGLLARALQDAAVDAGELNAALQARGEHSASRNPVFAPELVQWLQDALLLA
                NLELGRSQIDQAALILALLRNPLRYAGSRYQPLLARLDIERLKDFALSQPLEGAPSGK
                PGGGESLLQRFTHNLTQQARDGKLDPVLCRDGAIRQMVDILARRRKNNPIVVGEAGVG
                KTAIVEGLASRIAAGEVPPALRDVELLSLDMGLLQAGASVKGEFERRLKGVIDEVKAA
                PRPTILFIDEAHTLIGAGGQAGGSDAANLLKPALARGELRTIAATTWAEYKKYFEKDP
                ALARRFQPVQLHEPSVAEAVTILRGLARVYEESHGIYLRDDAVVAAAELSARYLAGRQ
                LPDKAVDVLDTACARVRISLAAAPESLERLRGELAEGERQRQALRRDAEAGLPIDHDA
                LQALEERLAQAESERHVQEDQWSRQRELAERLLALRQQLARAREANREEQGSEADAEA
                FPADEHGGVEALEKALHETRQALADAQAQGRLVSFEVCPRLVAEVIGAWTGVPVEQLA
                REHNAKVMGFADDLRTRIRGQEQAVSALDRAMRATAAGLNKPDAPVGVFLLVGPSGVG
                KTETALALADLLYGGERFITTINMSEFQEKHTVSRLIGAPPGYVGYGEGGMLTEAVRQ
                KPYSVILLDEVEKADPDVLNLFYQIFDKGVANDGEGREIDFRNTLILMTSNLASEHIA
                ALCQDGARPSAELLEETIRPALSRHFKPALLARMRVVPYYPVGGPVLRELIEIKLQRL
                GERLARRQLAFDYSQALVGHLAERCTQSDSGARLIDHLLDTHLLPLVADRLLEAMARD
                ERLERVHATLGEHGRVLCEFA"
gene            5505. .7016
                /gene="PA1663"
CDS             5505. .7016
                /gene="PA1663"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG05052.1"
                /db_xref="GI:9947634"
                /translation="MFSRVPQPLAFAEALLAQFGALSRETDITALLGGFARAVGELSG
                CELSQLYLLDQTHTALTLNAECRDGLLQALEPSSLPADYTGEQLLQFALCQNRVVSLV
                DLPSRLHETGFLPAGEQPWLSLLCVPLVSPEKTVEGMVVCATRRPLDLQGFAHPLGQL
                GSFVLAQLHLLQRLSQPSGEPQGPSSAPVVSGYGLIGRSAAMRETCRLISKVLHNPYT
                VLLQGETGTGKEVVARAIHEYGPRRAQAFVVQNCAAFPENLLESELFGYRKGAFTGAE
                RDRPGLFDAADGGTLLLDEIGDMPLSLQAKLLRVLQEGEIRPLGGNDTHKVDVRIIAA
                THRDLALLVRDGKFREDLYYRLAQFPIELPPLRQREGDALELARHFAERACVFLRRAP
                LGWSAEALELLDGYGFPGNVRELKGLVECAVLLCEGEMLLAEHFSLRAPPEAEAPLKL
                REHLEQIERDLLLACLRRNAGNQTLAARELGLPRRTLIYRMGRLNIHFGDPRT"
gene            7105. .7245
                /gene="PA1664"
CDS             7105. .7245
                /gene="PA1664"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05053.1"
                /db_xref="GI:9947635"
                /translation="MPVRHWPAVLLALAVLCGLAGCSGNYKFNDGDYRPLGDPQAVNR
                GK"
gene            7263. .8456
                /gene="PA1665"
CDS             7263. .8456
                /gene="PA1665"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05054.1"
                /db_xref="GI:9947636"
                /translation="MELVLEMLSAKQFVPTELCSKTFGRNGGLIGRGGECDWAIPDRK
                RHLSKQHARVSYRNGAFYLTDTSSNGIRAGNGSRLPHGEPQRIEQDSVFLLGDFEIRA
                RLLREPDELAMDVGRPQPAGSIIPDDAFLALDPLQALDQDGADAYELDDLSAIAQPSH
                EAGARADYARIDMESLVVPELVPAPAAPQEPVESPPPERDEAFWQRFGQALGVDLEGL
                DGAAREALAIDAAGLLRQCIGGLQQSLRTRSELKNELRLSLSTLRDTGKNPLRFSADA
                GEALGHLLRDGKPGQLSGEQAVARSFRDLQAHQVALLGASRAAVRATLEHFSPQQLAL
                RFERDGRRPLLATSGSRWRAYGRYHQSLCQDDDWSERLLARDFAQAYEEQVRLISTLY
                TEHQG"
gene            8462. .8968
                /gene="PA1666"
CDS             8462. .8968
                /gene="PA1666"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05055.1"
                /db_xref="GI:9947637"
                /translation="MSRRPIPPARALAMLAVLALLAGCSTLSPYSRLTKLDLALSAGE
                RVNPDLNGRPSPVVVRLFELKHPVAFENADFFSLYERPKETLDPDLVTSEELELRPGE
                RVELKLSVEEGSRYVGVLAAYRDLGEASWRYVVPVTAKELTRVELKLGQKGILNAGEL
                LGKAGDAR"
gene            8965. .10296
                /gene="PA1667"
CDS             8965. .10296
                /gene="PA1667"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05056.1"
                /db_xref="GI:9947638"
                /translation="MNVHKIVWQEGMLLRPQHLQHNDRYYDNQMKARTQLLGSYNWGF
                LALELDRQFLNMGKLVISQASGVLPDGSLFELGGNTEPLALDVPPNTSNVPIYLALPL
                VTGNHIEARRPEQAEVLARYTAYDLEVADSNAGEDSMVQVSCARPDFRLLLGEQQSDQ
                AYVKLKVAVVLDSTPDGVISLDPDFVPSYLHAHASAYLLSCLKEVISMLGHRGDTIAE
                RIRSNGKVGGAEVGDFMMLQLINRTELLLRHYLGLEQVHPEQLYRVLLTMLGDLATFG
                SESKRPRLDSRYQHSDQGASFRRLMEAIRQVLSMVLEQHAIELPLQARQYGILVSPLH
                DHKLLGSASFVLAASANCESEELRQRLPAHLKVGAVEHIRQLVNLHLPGIRVKPLPVA
                PRQIPFHSNKTYFILEPNAEDMAQLERSGGFAFHVSGEFSELELKFWAIRN"
gene            10299. .11168
                /gene="PA1668"
CDS             10299. .11168
                /gene="PA1668"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05057.1"
                /db_xref="GI:9947639"
                /translation="MIKEREYHEDEKTVLLDHQGQVPAQGPLTDFAEPPRFEQLEERM
                IYAARLRPGEAFNVSLNPLVAAASELLSEVVRLKHSTVLENLAALKERLTSAMKLFEF
                RAVQDGVENSQVMAARYVLCTVVDEAVVTTPWGNESEWSKISLLSSFHNETFGGEKFF
                QLLDRLSKNPVKHLPMLELMYLCLALGFEGKYRVMARGVLDLDGISDALYRQIRQLRG
                DVPRELSPHWQGLSDQRRGLVRIVPWWMVALFTLVCLVVMYSGFAWVLGEQRDTVLQP
                YQSLDAAAGQPGS"
gene            11184. .14711
                /gene="PA1669"
CDS             11184. .14711
                /gene="PA1669"
                /codon_start=1
                /transl_table=11
```

```
                     /product="hypothetical protein"
                     /protein_id="AAG05058.1"
                     /db_xref="GI:9947640"
                     /translation="MKNFFKKVGAFLRKTWVWTLLLVLLAALLVWFVGPLLAVADYKF
                     WESPTSRLLSISLMFLVWGLAMVFVSWRAGVRRKAVEDSDEGQERLRRDGLIEDEQKE
                     LRSRFKDALRTLKTSSLYRGRSERWRNELPWYLLIGPQGSGKTSLLDFSGLEFPINRI

  Query Match             73.3%;  Score 17.6;  DB 1;  Length 16985;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           |||||||||||  ||  ||  ||  |||
Db   15589 GGTAGCGCCCGGAAAGAACGTCCG 15566


RESULT   9
AC004396
LOCUS       AC004396               43903 bp    DNA     linear   HTG 13-MAR-2000
DEFINITION  Pseudomonas sp. chromosome Genomic clone cos3b, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC004396
VERSION     AC004396.12  GI:7230801
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Pseudomonas sp.
  ORGANISM  Pseudomonas sp.
            Bacteria; Proteobacteria.
REFERENCE   1  (bases 1 to 43903)
  AUTHORS   Roe,B.A.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43903)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:7229744.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15232: contig of 15232 bp in length
            *    15233    15352: gap of unknown length
            *    15353    43903: contig of 28551 bp in length.
FEATURES             Location/Qualifiers
     source          1. .43903
                     /organism="Pseudomonas sp."
                     /db_xref="taxon:306"
                     /chromosome="Genomic"
                     /clone="cos3b"
BASE COUNT     8111 a  13520 c  13821 g   8330 t    121 others
ORIGIN


  Query Match             73.3%;  Score 17.6;  DB 2;  Length 43903;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           ||||  ||||||||  ||  |||||  |
Db   37617 GGTAACGCCCGTAGCGGTCGGCTG 37640


RESULT  10
AX127153
LOCUS       AX127153               309400 bp    DNA     linear   PAT 11-MAY-2001
DEFINITION  Sequence 7069 from Patent EP1108790.
ACCESSION   AX127153 AX114121
VERSION     AX127153.1  GI:14041141
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 309400)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 7069 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
     source          1. .309400
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
                     /note="Seq 1 to long (3.309.400) split in 11, seq 7069
                     3.000.001 3.309.400 309.400"
BASE COUNT    70133 a  86477 c  83115 g  69675 t
ORIGIN


  Query Match             73.3%;  Score 17.6;  DB 6;  Length 309400;
  Best Local Similarity   83.3%;  Pred. No. 9.2e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           |||  |||||||  |||  ||||||  |
Db   70454 GGTTGCGCCCGGAACAATCGGCTG 70477


RESULT  11
AE009268/c
LOCUS       AE009268               11208 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 38 of
            187 of the complete sequence.
ACCESSION   AE009268 AE008689
VERSION     AE009268.1  GI:17741765
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (U. Washington).
  ORGANISM  Agrobacterium tumefaciens str. C58 (U. Washington)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 11208)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     The Genome of the Natural Genetic Engineer Agrobacterium
            tumefaciens C58
  JOURNAL   Science 294 (5550), 2317-2323 (2001)
  PUBMED    11743193
REFERENCE   2  (bases 1 to 11208)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
```

```
          98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .11208
                     /organism="Agrobacterium tumefaciens str. C58 (U.
                     Washington)"
                     /strain="C58"
                     /db_xref="taxon:180835"
     gene            complement(95. .1672)
                     /gene="Atu3375"
     CDS             complement(95. .1672)
                     /gene="Atu3375"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, substrate binding protein
                     [peptide]"
                     /protein_id="AAL44188.1"
                     /db_xref="GI:17741766"
                     /translation="MTINDNKEGRTVKKLLLAGVALLVTANFAAARDITIAQSSDLRS
                     SNPGVNRDGNTDGVILHIVEGLVGYNNAGEVKPLLAESVEVAPDGLAYTFKLRKDVKF
                     HNGKPLTAEDVVWNWTRYLKPETKWACLNDFAEGGAAHVTAVKATDASTVQITLEKPS
                     GVFLGLMSRPECGYTGMISPESVAADGSFVKPIGTGPFQWDEWKKGEYIRLAKFADYI
                     SPANDGKPDGMVGSKRPLADGIKFMVIPDASTVKAGLESGVLDTAEISPDLIPEFKTN
                     DKMQLIVSRNNGKNLFYIQTRDKVLSNPGVRRAMAMALDLDELVAAASNGTGEANGSM
                     VSQDSVYYDETQKRRPAYDVEAAKKELEAAGYKGEPISIIANKRGNVPSFPAAVMAQA
                     MMQQAGLNVQIEVLDYATQVDRRRSGNYQVISQSVAPRLDPALMYAFYVGNKDKNASL
                     MWDDPKAIELMKAAYVEADQAKRQRVFDEFHELMLKEMPGIFLYDMVDVWGATKKLKG
                     QPVWQSNARLWEVSVDN"
     gene            complement(1669. .2517)
                     /gene="carB"
                     /note="Atu3376"
     CDS             complement(1669. .2517)
                     /gene="carB"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAL44189.1"
                     /db_xref="GI:17741767"
                     /translation="MTMTVAEQLNAICDVQPFVTRFMVRNLLTGEEIGRGENEETPSG
                     STRKTSIMMAVLKAASEGRIDLDQPVTYEKRLAEEVASGMFRYMTPGIVISLRDAIAG
                     MMVLSDNVCTKMVLERLTLEEVDGYCKSLGMANTHHRFLIPPLALAADHPLKTVTTTS
                     AADQVLLLQTILDAQSSPEAQAKLGCSPQLCEWAMQTLKNQILRYGIRSRLPFEAMVA
                     SKSGRGKRGRMDAGIVYRNGAPSFIITTYTDEVPQVMPDGTPGYTIALETIGRLAQAC
                     WDGFKQ"
     gene            complement(2525. .4159)
                     /gene="Atu3377"
     CDS             complement(2525. .4159)
                     /gene="Atu3377"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, nucleotide binding/ATPase
                     protein [peptide]"
                     /protein_id="AAL44190.1"
                     /db_xref="GI:17741768"
                     /translation="MTDNTLLTVRGLSLEVAGSGHRVVKDVSFDIAPGEIFGIVGESG
                     SGKTLATRALISLLPPAIAVTGGSVIYKGQDVMSLPARELRRLRGAEIGVVFQEPMTS
                     LNPSMTIGRQLEEGLILHTKLSPEERREKILEMLRRVGIRDPEGALAAYPHEFSGGMR
                     QRIMLASVMLLKPALLIADEPTTALDAVIQRDVMELMVELTRAEGTAVLLISHDLPMV
                     ARYTSRIVVMEKGAIVEQGRTEDLLDAPRHPYTKKLLSSLPFRGEPRSIDTSKAPMVS
                     ARGIVVDYAGRKSLLKKAKPKRALHGVSIDIHEGEVVALVGGSGSGKTTLGRTIAGLV
                     KESEGHIRFQGRERNEDWNDYRLNCQMVFQDPYSSLDPRMTIQALVEEALRLVPGLDQ
                     AAKRKRALETLEEVGLGVDYAGRYPHELSGGQRQRVAIARAIARRPRFLIADEPVSAL
                     DVTVRAQVLDLFSDLQKRYGFSCLFISHDLGVVEQVADRVVVMQDGRIIEEGDRDTIF
                     DSPREAYTRRLLSAIPALDQNEKGGVTLKWRLEDEA"
     gene            complement(4156. .4974)
                     /gene="Atu3378"
     CDS             complement(4156. .4974)
                     /gene="Atu3378"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, membrane spanning protein
                     [oligopeptide]"
                     /protein_id="AAL44191.1"
                     /db_xref="GI:17741769"
                     /translation="MKKFTLNGLIGGILIALLLITAITGLFWTPYDPMKLGFASRLAA
                     PGAGHLLGTDEFGRDVLSRLMVGARASVWIGFLTVSFATICGTLIGLVSGYARGWVDG
                     VIMAINNALLAFPGILLALGLLAVFGANQYGIIFALGIAYTPSMARVVRGAVLSLRER
                     EFIEASLVMGNGEIYTMFRHILPNCIAPITVLATSMFGWAILSESALSFLGLGVPPPA
                     PTWGNMLAAGRPFIQQAVWLGLFPGLCIALTLLGINLLGDALRDRLDPRMRGLK"
     gene            complement(4971. .5915)
                     /gene="dppB"
                     /note="Atu3379"
     CDS             complement(4971. .5915)
                     /gene="dppB"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, membrane spanning protein
                     [peptide]"
                     /protein_id="AAL44192.1"
                     /db_xref="GI:17741770"
                     /translation="MIRFILSRLLMALPTIAFVSITVFALIRFIPGDPAALMLGDMAQ
                     PEQIEAMRTELGLDKSMLQQFVIWAGNVVQGDFGRSIVNDEAVLPLVVSRFLVSAEIV
                     VVAVLLASLIAVPAGVIAAWKQNSLTDLALVGTATLLLSIPTFWLGLLLLFFGLKLG
                     WLPVLGYVSISDNLVGGMLYLVLPVMTLVIHEMGVLIRMARASTLEVLRLDYITHARA
                     KGLSESAVLWRHAFKNAFGPTWTMIGLILGNLLGGIAVIETVFTIPGLGRLMVDSIFA
                     RDYPVIQGCLLFVSLSYVLVNLFVDLLYPLFDPRVVAE"
     gene            5902. .6120
                     /gene="Atu3380"
     CDS             5902. .6120
                     /gene="Atu3380"
                     /note="putative; ORF located using Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL44193.1"
                     /db_xref="GI:17741771"
                     /translation="MNLIIGSPRLMAGNDIPPLSVSLLVTMTDREISRNIMHKRHIAM
                     QNDGVFHPDDDNEISRPKSRFQSIFCES"
     gene            complement(6096. .7076)
                     /gene="Atu3381"
     CDS             complement(6096. .7076)
                     /gene="Atu3381"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, LysR family"
                     /protein_id="AAL44194.1"
                     /db_xref="GI:17741772"
                     /translation="MTRSTIGGKRFEPIPARFSSIGAGLMNIKQLEVLRTLLSTGSTI
                     ATAKAMGLSQSGISRLLAQLEADLSLTLFARDKGRLIPTPEAALLARDAENVLLAVDR
                     MSGHAEDLRNGAAGPEIVRIGLPSSMWENFAPAMLIDYVRDFPGVRIETFFETTTAIN
                     KLVGERVIDLGFLRMEGEIGPGIDVERVATGKSVCVVREDHPLADLQEITVKDLRNIP
                     LILIGRQRPNRMALDQVFKKAGVKPMVKIETHTNSSACAYVAHGLGVTIISSFYANLY
                     RHLPVVARPFVPQATQEFGLARASGAPLSIAAQALSNALKREIQLSQKID"
     gene            complement(7210. .8523)
                     /gene="prsE"
                     /note="Atu3382"
     CDS             complement(7210. .8523)
                     /gene="prsE"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="secretion protein, HlyD family"
                     /protein_id="AAL44195.1"
                     /db_xref="GI:17741773"
```

```
                     /translation="MPNIDQRTQVSRSIRKHLLAGLAAGAALLAGVGGWAATTNLAGA
                     VVASGHLVVDSYTKKVQHPKGGVVGEILVNEGERVKAGDVVMRLDATQTRANLAIVTK
                     RLDELGARMARLEAERDDLAELIFPQALIARRDDPDVASAMRSETKLFEFRKSYREGR
                     KAQLAERITQFEHEIEGLKAQEVAYDNGLAVLEAEITSQKSLREQGVVSVQRLNSLQT
                     QAATFGGERGEKIAYQAQTAGRITETKLQILQIDQELRTEVGRELREIQAQMGEFVER
                     KVAAEDELKRIDIVAPQSGMVHEMAVHTVGGVVTPADPIMLIVPDGDELALEVQIVPK
                     DIDQLQVGQKAMLRMTAFNQRVTPELEGHVSRIAADITTDQRTGLSYYLARISVPVSE
                     REKLNNAPLVPGMPAEAFIQTSERTALSYIAKPLTDQISRAFREE"
     gene            complement(8510. .10357)
                     /gene="prsD"
                     /note="Atu3383"
     CDS             complement(8510. .10357)
                     /gene="prsD"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"

  Query Match             71.7%;  Score 17.2;  DB 1;  Length 11208;
  Best Local Similarity   86.4%;  Pred. No. 1.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 tagcgcccgtaacgatcggccg 24
           | |||| || ||||||||||||
Db    1727 TCGCGCTCGAAACGATCGGCCG 1706


RESULT  12
AE008344
LOCUS       AE008344               12819 bp    DNA     linear   BCT 18-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 148
            of 187 of the complete sequence.
ACCESSION   AE008344 AE007870
VERSION     AE008344.1  GI:15159976
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon).
  ORGANISM  Agrobacterium tumefaciens str. C58 (Cereon)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 12819)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 12819)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Approximately 800 bp of telomeric sequence missing from the left
            end of the chromosome and 200 bp missing from the right end.
FEATURES             Location/Qualifiers
     source          1. .12819
                     /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
                     /strain="C58"
                     /db_xref="taxon:181661"
     gene            83. .838
                     /gene="AGR_L_2877"
     CDS             83. .838
                     /gene="AGR_L_2877"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2877p"
                     /protein_id="AAK90006.1"
                     /db_xref="GI:15159977"
                     /translation="MATTIQLTQSSSDLIGYLNGWTSGFGSTYGAFYDAQTSSLATTT
                     IDPNTTYEAWGDGSSGGKGIVMSGALQYSQGNLTGSVDSIVLGTGYSQSTSGIAVSQQ
                     ELLIDPDSAYSLAGARDLLDLAVYQLARFGSLAGFYDYFAATGTVIEDTAASNTLTGF
                     AGADTFVFSGGNDVVQAGPTGTYGYQDGTDTLDVSAWGATSVDDLLWYNDNGNAVILS
                     ATDTSVSITLNGVDANVLDASDFIFASALALAA"
     gene            853. .2700
                     /gene="AGR_L_2880"
     CDS             853. .2700
                     /gene="AGR_L_2880"
                     /note="(AF141932) rhiziobicin secretion protein RspD"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2880p"
                     /protein_id="AAK90007.1"
                     /db_xref="GI:15159978"
                     /translation="MGFQNHQQTPPQSLTAAVVSPLKSTFLGVALVSGVVNILALTSP
                     LFMLQVYDRVLASGSLPTLVGLALLALGLYGFQCLLDIIRARVLIRIGEDFDMRYSGR
                     VHDAVVRLPLVNRMPGDGLQPLRDLDNVRGFLSGTGPTAFFDLPWMPLYLGLCFLFHV
                     WIGVTALAGAVLLVSLTILTNIFSQKPIRDTMVENMARNRQLEASRRNAEVVQAMGLG
                     RRLGKRWQSSNEAYLAANRKAGDVAGGLGNLAKSLRVVLQSAILAVGAWLVIRQEASG
                     GVMIASSIMMGRALAPVDLAISNWKSFVAARQSWARLRELFAQMPANAAAMALPKPEK
                     ELRVENVTIVPPGERKPTVTGLGFVVTAGNALGIIGPSGSGKSTLARILTGAWMPAAG
                     KVRLDGASFDQWDRETLGRHIGYLPQGVELFDGTIGENISRFDDNPDPEAIIAAARAA
                     GAHELILRFEKGYDSDIGEAGSTLSAGQRQRIGLARALYGDPFIVVLDEPNANLDAEG
                     EAAVVKAITSVKARGGIAVVVAHRPSAIGAVDYILMMEDGRMKAFGPRDEVLSKVLRV
                     PQAQAAKGFAASAQTVSPLRVVANVQPKPALVDDVQDEESQEGDADAKH"
     gene            2576. .4000
                     /gene="AGR_L_2881"
     CDS             2576. .4000
                     /gene="AGR_L_2881"
                     /note="(AF141932) rhizobiocin secretion protein RspE"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2881p"
                     /protein_id="AAK90008.1"
                     /db_xref="GI:15159979"
                     /translation="MPHPPRPSRRYVSLPMSNPNPRWWTMFRMRRARKEMRMPNIDQR
                     TQVSRSIRKHLLAGLAAGAALLAGVGGWAATTNLAGAVVASGHLVVDSYTKKVQHPKG
                     GVVGEILVNEGERVKAGDVVMRLDATQTRANLAIVTKRLDELGARMARLEAERDDLAE
                     LIFPQALIARRDDPDVASAMRSETKLFEFRKSYREGRKAQLAERITQFEHEIEGLKAQ
                     EVAYDNGLAVLEAEITSQKSLREQGVVSVQRLNSLQTQAATFGGERGEKIAYQAQTAG
                     RITETKLQILQIDQELRTEVGRELREIQAQMGEFVERKVAAEDELKRIDIVAPQSGMV
                     HEMAVHTVGGVVTPADPIMLIVPDGDELALEVQIVPKDIDQLQVGQKAMLRMTAFNQR
                     VTPELEGHVSRIAADITTDQRTGLSYYLARISVPVSEREKLNNAPLVPGMPAEAFIQT
                     SERTALSYIAKPLTDQISRAFREE"
     gene            4134. .5114
                     /gene="AGR_L_2884"
     CDS             4134. .5114
                     /gene="AGR_L_2884"
                     /note="nocR protein - Agrobacterium tumefaciens plasmid
                     pTiC58  (fragment)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2884p"
                     /protein_id="AAK90009.1"
                     /db_xref="GI:15159980"
                     /translation="MTRSTIGGKRFEPIPARFSSIGAGLMNIKQLEVLRTLLSTGSTI
                     ATAKAMGLSQSGISRLLAQLEADLSLTLFARDKGRLIPTPEAALLARDAENVLLAVDR
                     MSGHAEDLRNGAAGPEIVRIGLPSSMWENFAPAMLIDYVRDFPGVRIETFFETTTAIN
                     KLVGERVIDLGFLRMEGEIGPGIDVERVATGKSVCVVREDHPLADLQEITVKDLRNIP
                     LILIGRQRPNRMALDQVFKKAGVKPMVKIETHTNSSACAYVAHGLGVTIISSFYANLY
                     RHLPVVARPFVPQATQEFGLARASGAPLSIAAQALSNALKREIQLSQKID"
     gene            5295. .6239
                     /gene="AGR_L_2886"
     CDS             5295. .6239
                     /gene="AGR_L_2886"
                     /note="PROBABLE PEPTIDE ABC TRANSPORTER PERMEASE PROTEIN
                     Y4TP"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2886p"
                     /protein_id="AAK90010.1"
                     /db_xref="GI:15159981"
                     /translation="MIRFILSRLLMALPTIAFVSITVFALIRFIPGDPAALMLGDMAQ
                     PEQIEAMRTELGLDKSMLQQFVIWAGNVVQGDFGRSIVNDEAVLPLVVSRFLVSAEIV
                     VVAVLLASLIAVPAGVIAAWKQNSLTDLALVGTATLLLSIPTFWLGLLLLLFFGLKLG
                     WLPVLGYVSISDNLVGGMLYLVLPVMTLVIHEMGVLIRMARASTLEVLRLDYITHARA
                     KGLSESAVLWRHAFKNAFGPTWTMIGLILGNLLGGIAVIETVFTIPGLGRLMVDSIFA
                     RDYPVIQGCLLFVSLSYVLVNLFVDLLYPLFDPRVVAE"
     gene            6236. .7054
                     /gene="AGR_L_2887"
```

```
CDS             6236. .7054
                /gene="AGR_L_2887"
                /note="DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC"
                /codon_start=1
                /transl_table=11
                /product="AGR_L_2887p"
                /protein_id="AAK90011.1"
                /db_xref="GI:15159982"
                /translation="MKKFTLNGLIGGILIALLLITAITGLFWTPYDPMKLGFASRLAA
                PGAGHLLGTDEFGRDVLSRLMVGARASVWIGFLTVSFATICGTLIGLVSGYARGWVDG
                VIMAINNALLAFPGILLALGLLAVFGANQYGIIFALGIAYTPSMARVVRGAVLSLRER
                EFIEASLVMGNGEIYTMFRHILPNCIAPITVLATSMFGWAILSESALSFLGLGVPPPA
                PTWGNMLAAGRPFIQQAVWLGLFPGLCIALTLLGINLLGDALRDRLDPRMRGLK"
gene            7051. .8685
                /gene="AGR_L_2889"
CDS             7051. .8685
                /gene="AGR_L_2889"
                /note="probable ATP-binding component of ABC transporter
                PA2061 (imported) - Pseudomonas aeruginosa (strain PAO1)"
                /codon_start=1
                /transl_table=11
                /product="AGR_L_2889p"
                /protein_id="AAK90012.1"
                /db_xref="GI:15159983"
                /translation="MTDNTLLTVRGLSLEVAGSGHRVVKDVSFDIAPGEIFGIVGESG
                SGKTLATRALISLLPPAIAVTGGSVIYKGQDVMSLPARELRRLRGAEIGVVFQEPMTS
                LNPSMTIGRQLEEGLILHTKLSPEERREKILEMLRRVGIRDPEGALAAYPHEFSGGMR
                QRIMLASVMLLKPALLIADEPTTALDAVIQRDVMELMVELTRAEGTAVLLISHDLPMV
                ARYTSRIVVMEKGAIVEQGRTEDLLDAPRHPYTKKLLSSLPFRGEPRSIDTSKAPMVS
                ARGIVVDYAGRKSLLKKAKPKRALHGVSIDIHEGEVVALVGGSGSGKTTLGRTIAGLV
                KESEGHIRFQGRERNEDWNDYRLNCQMVFQDPYSSLDPRMTIQALVEEALRLVPGLDQ
                AAKRKRALETLEEVGLGVDYAGRYPHELSGGQRQRVAIARAIARRPRFLIADEPVSAL
                DVTVRAQVLDLFSDLQKRYGFSCLFISHDLGVVEQVADRVVVMQDGRIIEEGDRDTIF
                DSPREAYTRRLLSAIPALDQNEKGGVTLKWRLEDEA"
gene            8675. .9541
                /gene="AGR_L_2891"
CDS             8675. .9541
                /gene="AGR_L_2891"
                /note="probable beta-lactamase - Deinococcus radiodurans
                (strain R1)"
                /codon_start=1
                /transl_table=11
                /product="AGR_L_2891p"
                /protein_id="AAK90013.1"
                /db_xref="GI:15159984"
                /translation="MKLNRKMTMTVAEQLNAICDVQPFVTRFMVRNLLTGEEIGRGEN
                EETPSGSTRKTSIMMAVLKAASEGRIDLDQPVTYEKRLAEEVASGMFRYMTPGIVISL
                RDAIAGMMVLSDNVCTKMVLERLTLEEVDGYCKSLGMANTHHRFLIPPLALAADHPLK
                TVTTTSAADQVLLLQTILDAQSSPEAQAKLGCSPQLCEWAMQTLKNQILRYGIRSRLP
                FEAMVASKSGRGKRGRMDAGIVYRNGAPSFIITTYTDEVPQVMPDGTPGYTIALETIG
                RLAQACWDGFKQ"
gene            9538. .11115
                /gene="AGR_L_2894"
CDS             9538. .11115
                /gene="AGR_L_2894"
                /note="agaA"
                /codon_start=1
                /transl_table=11
                /product="AGR_L_2894p"
                /protein_id="AAK90014.1"
                /db_xref="GI:15159985"
                /translation="MTINDNKEGRTVKKLLLAGVALLVTANFAAARDITIAQSSDLRS
                SNPGVNRDGNTDGVILHIVEGLVGYNNAGEVKPLLAESVEVAPDGLAYTFKLRKDVKF
                HNGKPLTAEDVVWNWTRYLKPETKWACLNDFAEGGAAHVTAVKATDASTVQITLEKPS
                GVFLGLMSRPECGYTGMISPESVAADGSFVKPIGTGPFQWDEWKKGEYIRLAKFADYI
                SPANDGKPDGMVGSKRPLADGIKFMVIPDASTVKAGLESGVLDTAEISPDLIPEFKTN
                DKMQLIVSRNNGKNLFYIQTRDKVLSNPGVRRAMAMALDLDELVAAASNGTGEANGSM
                VSQDSVYYDETQKRRPAYDVEAAKKELEAAGYKGEPISIIANKRGNVPSFPAAVMAQA
                MMQQAGLNVQIEVLDYATQVDRRRSGNYQVISQSVAPRLDPALMYAFYVGNKDKNASL
                MWDDPKAIELMKAAYVEADQAKRQRVFDEFHELMLKEMPGIFLYDMVDVWGATKKLKG
                QPVWQSNARLWEVSVDN"
gene            11027. .12685
                /gene="AGR_L_2896"
CDS             11027. .12685
                /gene="AGR_L_2896"
                /note="AMINOBENZOYL-GLUTAMATE UTILIZATION PROTEIN B"
                /codon_start=1
                /transl_table=11
                /product="AGR_L_2896p"
                /protein_id="AAK90015.1"
                /db_xref="GI:15159986"

  Query Match             71.7%;  Score 17.2;  DB 1;  Length 12819;
  Best Local Similarity   86.4%;  Pred. No. 1.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 tagcgcccgtaacgatcggccg 24
          | ||||  || ||||||||||||
Db   9483 TCGCGCTCGAAACGATCGGCCG 9504


RESULT  13
AE000030
LOCUS       AE000030               11802 bp    DNA     linear   BCT 24-NOV-2000
DEFINITION  Mycoplasma pneumoniae M129 section 50 of 63 of the complete genome.
ACCESSION   AE000030 U00089
VERSION     AE000030.2  GI:11379517
KEYWORDS    .
SOURCE      Mycoplasma pneumoniae.
  ORGANISM  Mycoplasma pneumoniae
            Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
            Mycoplasmataceae; Mycoplasma.
REFERENCE   1  (bases 1 to 11802)
  AUTHORS   Himmelreich,R., Hilbert,H., Plagens,H., Pirkl,E., Li,B.C. and
            Herrmann,R.
  TITLE     Complete sequence analysis of the genome of the bacterium
            Mycoplasma pneumoniae
  JOURNAL   Nucleic Acids Res. 24 (22), 4420-4449 (1996)
  MEDLINE   97105885
REFERENCE   2  (bases 1 to 11802)
  AUTHORS   Dandekar,T., Huynen,M., Regula,J.T., Ueberle,B., Zimmermann,C.U.,
            Andrade,M.A., Doerks,T., Sanchez-Pulido,L., Snel,B., Suyama,M.,
            Yuan,Y.P., Herrmann,R. and Bork,P.
  TITLE     Re-annotating the mycoplasma pneumoniae genome sequence: adding
            value, function and reading frames
  JOURNAL   Nucleic Acids Res. 28 (17), 3278-3288 (2000)
  MEDLINE   20411492
REFERENCE   3  (bases 1 to 11802)
  AUTHORS   Himmelreich,R., Hilbert,H. and Li,B.-C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Zentrun fuer Molekulare Biologie
            Heidelberg, University Heidelberg, 69120 Heidelberg, Germany
REFERENCE   4  (bases 1 to 11802)
  AUTHORS   Suyama,M., Dandekar,T. and Herrmann,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2000) Zentrun fuer Molekulare Biologie
            Heidelberg, University Heidelberg, 69120 Heidelberg, Germany
COMMENT     On Nov 27, 2000 this sequence version replaced gi:1673996.
            This updated annotation replaces the old annotation from reference
            1. The old gene identifiers (MP numbers) according to the original
            publication by
            Himmelreich et al. (1996) are given as well as new gene numbering
            (MPN numbers) from the origin of replication. Annotation comments
            and further update data are at
            http://www.bork.embl-heidelberg.de/Annot/MP/.
FEATURES             Location/Qualifiers
     source          1. .11802
                     /organism="Mycoplasma pneumoniae"
                     /strain="M129"
                     /db_xref="taxon:2104"
                     /note="ATCC 29342"
     gene            complement(104. .3976)
                     /gene="rpoC"
                     /note="F04_orf1290"
     CDS             complement(104. .3976)
                     /gene="rpoC"
```

```
                     /note="MPN515(new), 327(Himmelreich et al., .1996)"
                     /codon_start=1
                     /transl_table=4
                     /product="DNA-directed RNA polymerase beta' chain"
                     /protein_id="AAB95975.1"
                     /db_xref="GI:1674001"
                     /translation="MTKRNKKNNKLYKNIKAIKLSIASNDTILNWSEGEVTKAETINY
                     KSLKPEPGGLFDEAIFGPVKDYECACGKFKKIKYRGVRCDRCGVWVTESIVRRERMGH
                     IALVSPVAHIWMSKELPSPSKISLVLNISYKEVEQVLYFVNYIVLDTGKIKDPKIMPF
                     KFKEVLDLAGKGSLTTRQKMRRVIGYIFRNLIKNRSSEDYRKGKIFYESLKNSSLPFS
                     LNDAFNYIKKYTGFRVGIGAEAILELLNKIDLNYEFSKLNDALRKAKKDSVEDAKVKK
                     ILRQLETISWFRNSKLHPKNMILHTVPVIPPDIRPIIQLDGAKFTTSDINNFYRRVII
                     RNDRLRRILEDGTVPAIVVNNEKRLLQESVDALFDNSSRHKPALSKDKRSLKSLTDRL
                     KGKQGLFRHNLLGKRVDYSGRSVIVVGPELKMYEVGIPALMILKLFKPFIIHGLINKF
                     DSNGNEIRPIASSIRQAEDMIKNQDDLIWGIVYDVIKDRPVLLNRAPTLHRLGIQAFE
                     PRIVDGKAIRLHPLVTTAFNADFDGDQMAVHVPLSENAVNEARAILLASKHILGLKDG
                     RPIVTPTQDMVLGNYYLTTERKGQTGEGIIFGTVHEARAAYEAGKVHLHAIVGISTKA
                     FPNKHFEAQGTLITTVGKIIFNDVLGDNIPYINEGEFDEHACPQKFIVPPSGDVRAAI
                     AAHQVLPAFGKKVISKLIDLLYTVVEFKDLPRILENIKALGFKYSTHSSTTVSVFDIP
                     KYSNKQQYFDEADQQVLKYKQFYNKGLLTDDERYKRVVKLWNGVKEKVSSEIQDLIKR
                     EEYRDNSIVVMADSGARGNISNFTQLFGMRGLMSKSFNYERNNQSKIIKDTIEVPIKH
                     SFLEGLTINEYFNSSYGARKGMTDTAMKTAKSGYMTRKLVDATHELIINHDDCGTRKG
                     IVVEAIVETKTRSLVESLFDRIVNRYTIGPILDPETKAEIVPANSLITQELAKQICAT
                     SIKQVLVRSVIYCERENGVCQYCFGVDLSTGKLVELGTAVGVIAAQSIGEPGTQLTMR
                     TFHTGGVSTENNLAQGFERLKQIFEVVAPKDYERCVISEVKGVVKSITTTQNAQEVLI
                     ESSVDERTYSIPFSAQLRVKVGDAVELGSKITEGSIDIRQLLRVAGIQRVRQYMIVEI
                     QKVYRIQGIEIADKYVEIIIRQLTSLLQVTDAGSSNLFVGQLVHSHHLNELNKSLLLS
                     GKMPVIAINQVFGIDEAASKSNSFLSAASFQDTKKILTDAAVKTQVDYLLGLKENVII
                     GGKIPAGTGFLTDEELAYLGAKTVQEEY"
gene                 complement(3980. .8155)
                     /gene="rpoB"
                     /note="G12_orf1391o"
CDS                  complement(3980. .8155)
                     /gene="rpoB"
                     /note="MPN516(new), 326(Himmelreich et al., 1996)"
                     /codon_start=1
                     /transl_table=4
                     /product="RNA polymerase beta subunit"
                     /protein_id="AAB95974.1"
                     /db_xref="GI:1674000"
                     /translation="MSQKPSFFQKKYSPTATRRYYGKIATDFVQPNLADIQIRSYQTF
                     LDHDLENLIAAYFPIKSPNDRYTINFKGLRRTAPERNEAQSRSESKTYEIGIYADLEL
                     IDSATGTIKKPRKSKKNSATSSVDGVFLTNLPLITRDGVFIVNGIEKFVIAQITRSPG
                     IYMLTKSQLKLSSSRKRVQEGYVCEVLPANGSVMLIYISNKKKIEDAFVQILLRDAVR
                     EGAKIFPITTLLKAFGMSGKEILKVFKNNEFITRSLEAEVYNAKDFLNNVDPEIKNLL
                     REFRDGKTDLRRKGIASDQKIRSLVSDYVLLEKEHKALSEAKPNDPKVGQLEADMDEL
                     MDKIITERAAKHIVHELSISLRGLENTDECPENSYHALLCSRFFRQRRYNLSAAGRYK
                     VSRKLRITERIYQKTLACDLHLKNGELLLKKGTLLVKEEIDKIKQAAQNNQIDFVQKI
                     KLTTDGSAVNLSPESLLYESLDVYVNNDNFDVSVPVVGIHNDNDLNKAITLSDFIASI
                     SYVINIPSAIGKYDDIDHLGNKRVKLINELISSRLESGITRMERFLKEKLTIADGVNR
                     GQQINEEGQVIEQAEKKELTIKSLINSKPIQIVIRDFFNTHQLTQFLDHQNPLSELSN
                     KRRISAMGPGGISREDPNLDIRDVHYSQYGRICPIETPEGMNIGLIMSLASFAKIDEN
                     GFLMAPYRKIKNGVITDEVEYLTALREDEHIIAEISSLVNIDENNKILDKEIIGRYRS
                     MQGLYDPSKIDYIDVAPHQVVSIGSSLIPFLENDDSARALMGTNMQRQAYPLIKPYAP
                     VVGTGQEYKIARDSGLTMLAPCSGTVKYVDNSKITIESDSGEQHTLDLIKFERSNQNT
                     CYNHVPLVEKGQRVTKDEVIADGPAVNKSELSLGQNVLVAFTTWNGYNYEDAIVISER
                     LVKDDVLTSLTINEYVAQCLSTKNGDEQITRDIPNVSDANKRYLDENGIIMVGAEVKE
                     GDVLVGKVSPKGQVEVSPEEKLFKAIFPESVQNVRDSSLKLPHGGDGIVSCVKRFSIA
                     NGNELNDGVIEMIKVYVVQKRKIQIGDKLAGRHGNKGVISKVVPVADMPHLEDGTPVD
                     ILLNPLGVPSRMNIGQIFEMHLGYAAHNLAKRMLISACFDDKKAQALSTEINQPQYKL
                     DRLITGLKAQITNRGLKDEQAALAQLNNGDIALVLKEIGMSFDDLHFKVATPIFQGVN
                     FQDLQDIMDEAGLKPAETHGKFKLIDGRTGLPFEKPISLGIMYIMKLNHMVDDKIHAR
                     AVGPYSKITQQPLGGKSQNGGQRFGEMEVWALEAYGAAYNLQELLTIKSDDVQGRNKA
                     YAAIVKGAAFPEPGIPESFKLLTKELQGLALSVSFIYDDNTQQDSNNVSILQADGEQD
                     DLFNDFEFDTEGY"
gene                 complement(8248. .8748)
                     /gene="G12_orf166a"
CDS                  complement(8248. .8748)
                     /gene="G12_orf166a"
                     /note="MPN517(new), 325(Himmelreich et al., 1996)"
                     /codon_start=1
                     /transl_table=4
                     /product="suggest reductase homologue"
                     /protein_id="AAB95973.1"
                     /db_xref="GI:1673999"
                     /translation="MSTKPLILLLANSKNSINRKFAKALEQQLNAELIELVDYQVDFY
                     CEDLEKDHFPEKIKSLVRKLHDHKTLIFVTPEHNGFVPAFAKNTIDWMTRDTQYGKNQ
                     FLKELDGIICCVTPAAKSGGKTVLELLTKFFSFSGLNVKGAVLVNGYHDGFDFQPFIT
                     DVQKLI"
gene                 complement(9841. .10887)
                     /gene="G12_orf348"
CDS                  complement(9841. .10887)
                     /gene="G12_orf348"
                     /note="MPN518(new), 324(Himmelreich et al., 1996)"
                     /codon_start=1
                     /transl_table=4
                     /product="conserved hypothetical protein"
                     /protein_id="AAB95972.1"
                     /db_xref="GI:1673998"
                     /translation="MAQTTINEQTLVNEQRLLAYDCFQNTNKVGLLSALEYLDFVNHL
                     VKSNQIKYLLVPSASQTVYQHVFFDHFRFLSFDNNLLSGFGLAIRLHKEAETVFMLVE
                     KTKQTETEIKKLLEAVKQFSGLKLVVLLVNSPFENKRTKLSYEIVKIIKKHGYQYKRY
                     CRKNFEKKFFNTWNHQKRKQKTKNNLFFVELNSQFAFGTSMANSNSFDLSFTSFKERF
                     NTENNLDIDVHFNVPHKYLLKNLESIKHHLQIDNNTIWKQAFINFAGQLLLYFENLQV
                     FAQEASTLQVNAIVIKQDAYFIVEVMKGYNYDNNCICLLEANKDQNVGLTTASTNTLI
                     GDIKQYEPCFFIEK"
gene                 complement(10890. .11708)
                     /gene="lip3"
                     /note="G12_orf272V"
CDS                  complement(10890. .11708)
                     /gene="lip3"
                     /note="MPN519(new), 323(Himmelreich et al., 1996)"
                     /codon_start=1
                     /transl_table=4
                     /product="triacylglycerol lipase (lip) 3"
                     /protein_id="AAB95971.1"
                     /db_xref="GI:1673997"
                     /translation="MRDKVNINTVDISQLEVFFQPAKKPAKQTIVFAHGFSVHHSYFK
                     SFSETLVDYDYYAPLWPGHNHHGFTDKELSPIHYAHLLVAWIEKQDLENIVLIGHSMG
                     GAVASYALQFLKPQRVEKLVLLAPLSYSNLLNYYKIKKAFKKKDEKLSYFKRMFQPKF
                     PDLQNDGQWQMELDKHTAMCKKLTFQIFKELSRLNSAYKTITIPAFLLLAQHDDFMPT
                     KATLNYFNRFLIKKGNLQSGVILHSQHQMFNSKHESFCKAMHDILKHNKLSKIY"
BASE COUNT     3208 a    2564 c    2322 g    3708 t
ORIGIN


  Query Match             70.0%;  Score 16.8;  DB 1;  Length 11802;
  Best Local Similarity   90.0%;  Pred. No. 2.5e+03;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcg 20
           || ||||||| |||||||||
Db    4169 GGCAGCGCCCTTAACGATCG 4188


RESULT  14
AE008205
LOCUS       AE008205               12850 bp    DNA     linear   BCT 18-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 9 of
            187 of the complete sequence.
ACCESSION   AE008205 AE007870
VERSION     AE008205.1  GI:15158363
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon).
  ORGANISM  Agrobacterium tumefaciens str. C58 (Cereon)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 12850)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 12850)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Approximately 800 bp of telomeric sequence missing from the left
            end of the chromosome and 200 bp missing from the right end.
FEATURES             Location/Qualifiers
     source          1. .12850
                     /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
                     /strain="C58"
                     /db_xref="taxon:181661"
     gene            166. .1398
                     /gene="AGR_L_149"
     CDS             166. .1398
                     /gene="AGR_L_149"
                     /note="(AL121855) putative membrane protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_149p"
                     /protein_id="AAK88641.1"
                     /db_xref="GI:15158364"
                     /translation="MRITLLGQFGSGNSGNDGSLEAMLLYLRRMRPDADLLCICASPT
                     VVSAKFNIRSMGVGGPPLTNPWMRQLDRLLGKVPSRLMLLVSCLFGFDKADLMIIPGT
                     GILDDFQEKWFGWPFVVFCWCLVARLRNTRIAFVSIGAGPMKSRLSRWFLRSAAQMAS
                     YRSYRDDFSLNYMKMIGVDVSRDHRYPDIAFDLPAPVPHRERSADGAMRIGVGVMDYR
                     GWRHNDPEANRIYQTYIHKLSQLVSWLIGEGHRITLLMGDVTDRVACADLMFLLNDML
                     SEDEMSQIEISSAATLRDIMLQMAAIDLAVVSRYHGVVCSLKLGNPTISIGYARKFDD
                     LMNDFDQEKFVRHIDTFEVEDVIELVKEILAGADAVRQRIAHANERFRRQLAEQEALL
                     GREFLLTDIVATPQGVTP"
     gene            1483. .3135
                     /gene="AGR_L_151"
     CDS             1483. .3135
                     /gene="AGR_L_151"
                     /note="GumJ protein - Xanthomonas campestris"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_151p"
                     /protein_id="AAK88642.1"
                     /db_xref="GI:15158365"
                     /translation="MPFEKHSPNRLHLRIRLASQWSDFQGSSPHSRFDATVDYRRRCR
                     SSSMHKVRRAFFMAMLEQHLGIAINFALIAIMSRLLSPAEVGFAVIGYGVTAVIFAFR
                     EFVSSDFLIQLDVVERDDINTSLTLLLLVSGILGVVLYLCGPYLALFYHQPGLQHYLV
                     LAIAAVMAESLGMPAFSMLRRRLAFGTIARVRTIGLVVMAVVTIFAARNGWGFVSFAL
                     GLLVGNSVASALAAYADPAQRSARLSLKSWRRMAEFGRYRGAATIVDKLYESLPQLVL
                     GFAMTPTAVAIYNRANTVCSIPDRIFLSAIFSFAFPALAAEVREGGDVRRSYLRALSY
                     ITVLYWPSLIMVAILADPIVRLALGDQWLAAVPITRILALASIFWFPMVLTTPLLIAL
                     NANREAFLGNFVCRSVSAIVLCSASFFGLTAVALSQFVTLPFQMLVAFYLVRRHAHFK
                     LPELFAAVAPSASVTLAAIAGPLAISAWMGFRFDYSLAQAGGAVLLSVTGWAAGLFLL
                     RHPLLSEIRSTLDAFTGRRRKVGKPAAPASIPDTVVSAAVEP"
     gene            3132. .4091
                     /gene="AGR_L_153"
     CDS             3132. .4091
                     /gene="AGR_L_153"
                     /note="glycosyl transferase (lgtD) RP446 - Rickettsia
                     prowazekii"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_153p"
                     /protein_id="AAK88643.1"
                     /db_xref="GI:15158366"
                     /translation="MSSIDIVIPNYNYGRFLKGCVESVLGQNIADMRILIIDNASTDD
                     SVAISRSLAAANPRIETSLRARNLGGHASFNAGIDWARADYFAIICADDVLSPGALAR
                     AMEALDNNPGAHMAFGRTAFFRDDAEIASGLHSDADGMRLHRGMDFIANICATGRSPV
                     QGPLAVTRTHIQQHAGHYRPALRHTDDVEMWMRLAVLGDIVELDAVQNFTRIHGANQS
                     AILDNVLAWNAEIEAALESFFGNEGALLADAKTLLRRGRASLSDRAYWCALSNLVRRR
                     PGAGALLAQAFRLRPASALLPPISYLMRRPDALHRIRAAIARR"
     gene            complement(4125. .6533)
                     /gene="AGR_L_153gl"
     CDS             complement(4125. .6533)
                     /gene="AGR_L_153gl"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_153glp"
                     /protein_id="AAK88644.1"
                     /db_xref="GI:15158367"
                     /translation="MAEGGDDSQKVPSAFEADTGPLPGSREAAHAKAFPSVGVGSSSS
                     GLNAKDQSMNGELVTSKEELQSLNEELTALNNQLHEALERQRAVANDLQNVLYSTDVP
                     TLFLDSSLRIRFFTPATRSVFHIVSSDVGRPLSDLSGFDGDGTLQTDAMATLHDGNTI
                     DREIGGSDGRWYVRRVKPYRGESKTVDGVVITFIDVSAQKQAAESREAAKKSAERATE
                     AKSRFLSAASHDLRQPLQTLKLLQGLLERSVGEGRARTFVQRMEETLSSMSGMLDSLL
                     DINQIEAGMVEPQVVNFAIGNLLGLLFKEFYYPARAAGLQLHVVPSSLAVRTDPRLFE
                     QIVRNLLSNALKYTTSGKILVGCRRRGTKVRFEVWDTGIGIPESQIEDVFQEYLQLDA
                     SAGTEDRGLGLGLSIVKRLSELLELRVAVRSRLHKGSVFMIEIDQAPAGEPLVPLRAP
                     VDGRPHHTTALSSASVLIIEDENGMRDLLKMGLEQSGYVVAATSTGAEALAIVRGLQF
                     KPDVILADYNFPTGGNGIAAIEDIRQALRRNVPALILTGDISSKALTTFAQHEIPYLH
                     KPVKLKDVVQVVETLVVRQTGERAEHISEPEDMEPMTGRLIEVVDDEKDIRDHLRALF
                     EGEGWTVATYVSAEEYIADYKPDRSGCLLVDAYLPGMVGMELLRLLKERQHRFPVIVI
                     TGHSDVRMAIDAMRYGAADFIEKPISYQELYRSVRSALDRSRDANEREERREAARATL
                     ATLTRKQRQILDRIVQGQANKIIAADLKLSQRTVENHRASIMRRTKSASLPALLRLVL
                     NAEE"
     gene            7175. .7963
                     /gene="AGR_L_157"
     CDS             7175. .7963
                     /gene="AGR_L_157"
                     /note="(AF076290) putative response regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_157p"
                     /protein_id="AAK88645.1"
                     /db_xref="GI:15158368"
                     /translation="MRLCQARQIRKHALARIGHTMLTSVPNNAYSPLSPNFHVIAHSK
                     PMETSVKKLAIIDDRALDRECLAQSLIAYGLDMEVELFSSLDDWRHAPNGHHSGILVN
                     VGRSDFCDDVSLHELREFIADYPYLPVVILAENRDLRQVMRAFDAGVRGYISSSIGLA
                     VCVGAISLALAGGAFVSAENLSDLRQLLVVAEDKERQRTAMFTQREVDVINALTQGKP
                     NKIIAYELNLRESTVKVHIRNIMKKVDAKNRTEVIFKISDLFKN"
     gene            8152. .9582
                     /gene="AGR_L_159"
     CDS             8152. .9582
                     /gene="AGR_L_159"
                     /note="(AL121855) putative aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_159p"
                     /protein_id="AAK88646.1"
                     /db_xref="GI:15158369"
                     /translation="MDIWTRDFAMQTFVENFENSDRLRQRARNIIPGGAHTYAKGDDQ
                     YPVLSPGFIERGYGCHVWDVDGNEYIEYGMGNRAVGLGHAYADVVEAARAELYRGCNF
                     TRPSKIEVDCASTFLDTVQTAEMVKFCKNGSDATSAAIRLARAVTGRDLIARCSDHPF
                     FSTDDWFIGTTEMNAGIPEAIQNLVLGFSYNDLASARSLFDRFKGRIAAFILEPSRGD
                     PPRDQFLHKLQKLCHENGALFILDEMITGFRWDAGGAQKTYDIVPDLSCFGKALGNGF
                     SISALAGKREFMERGGIDQTDAPRVFLLSTTHGAETHAMAAAIATMKVYTHEPVVPHL
                     VRQGNSLRAAISQIAIAHGLDSYFGIQGHSACLSYFTLDQTGKPSQAFRSLFLQETIR
                     RGLLAPSLVVSYTHEDEDIAKTLDAIDGALVVYRKALEYGVDKFLTGRPSDVVFRRFN
                     GNAQHQAGHGLVMADMPGLSESIRKP"
     gene            9634. .10770
                     /gene="AGR_L_161"
     CDS             9634. .10770
                     /gene="AGR_L_161"
                     /note="(AL121855) putative epimerase/dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_161p"
                     /protein_id="AAK88647.1"
                     /db_xref="GI:15158370"
                     /translation="MHRSRLKRYARRNRKWILRGSFMHVLVTGHRGYIGSVMVPILRK
                     AGHDVHGYDTEFYHRCAYVPGGPLPQVPGVCKDIRDITPADMEGFDAVIHLAALSNDP
                     LSNLNPEITYEINYRGSVRAAQAAKQAGVKRFIFASSCSNYGAAGEGLVNETAELKPV
                     SAYGTSKVLAEREIASLASPGFSPVYMRPATAYGLSPMLRFDIVLNNLTAWAVTEGLI
                     LLKSDGSPWRPIVHIEDISRAFLAALEAPTDAIHNQAFNVGQDAHNYQIREIAEIVAL
                     TVPGCRLQYASDASPDTRSYRVDFGKIARALPAFKPTWTAAAGAEQLYRAYKSAGLTM
                     AEFEGPHYQRIGQLQKLIGDKDIDLNLRRPMAAPRHGADIAAPV"
     gene            10793. .12127
                     /gene="AGR_L_42GM"
     CDS             10793. .12127
                     /gene="AGR_L_42GM"
                     /note="hypothetical protein"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="AGR_L_42GMp"
                     /protein_id="AAK88648.1"
                     /db_xref="GI:15158371"
                     /translation="MEPGIRLTGDAWRQEGIGQTIHDLASRLFPICRSITGKGVRDTI
                     DILKDYIDIEQRAVPSGTRVFDWTVPQEWIVRDAYVKNEAGRKVIDFQISNLHLMNYS
                     IPQRRVMPLGELQAHIHTLPAQPDLIPYKTAYYTDAWGFCMAHHDFLKLEEGDYEVVI
                     DTEKTSGNLFYGEYLHRGKNAKEVLLFAHICHPSLANDNCSGLALLTCLAAGLKSRST
                     YYSYRFVFAPGTIGSLSWLARNEEGISNIAHGLVVSCIGDGGGPNYKRSRRGDALIDR
                     IAACTRCGEDRKHLEIADFIPYGYDERQFCSPGFNLPVGLLQRSAFGTFPEYHTSADN
                     LDFIRPEHLASSFWMIMDMIDALEDNWTPLNLRPKGEPQLGKYGLFPAIGGHKTNDDR
                     TMAYLWILNLADGNNSLVDIAERSKMPFVDIASAADRLLEAGLVSNLGVQTR"
     gene            complement(12132. .12677)
                     /gene="AGR_L_166"
     CDS             complement(12132. .12677)
                     /gene="AGR_L_166"
                     /note="dTDP-6-deoxy-L-mannose-dehydrogenase -
                     Synechocystis sp.  (strain PCC 6803)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_166p"
                     /protein_id="AAK88649.1"
                     /db_xref="GI:15158372"
                     /translation="MHQERRMLFEKTKIEGLLVVNTIPVRDSRGSFARTFCKREFESN
                     ALVSDFSQHSLSLSKLKHTIRGLHFQRPPHQETKLVSCVKGSIWDVAVDLRRGSPTYL
                     QWVGTVLSADNGAQFYIPEGFAHGFQSLSDDVAVSYLISVPYAEASAAGIRHDEPAIG
                     VEWPAPASVISEKDLGWPAWQ"
BASE COUNT     2706 a   3851 c   3476 g   2817 t
ORIGIN


  Query Match           70.0%;  Score 16.8;  DB 1;  Length 12850;
  Best Local Similarity  90.0%;  Pred. No. 2.5e+03;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 tagcgcccgtaacgatcggc 22
          | ||||||| ||||||||||
Db  11469 TTGCGCCCGGAACGATCGGC 11488


RESULT 15
AE009408/c
LOCUS       AE009408               12952 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 178
            of 187 of the complete sequence.
ACCESSION   AE009408 AE008689
VERSION     AE009408.1  GI:17743302
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (U. Washington).
  ORGANISM  Agrobacterium tumefaciens str. C58 (U. Washington)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 12952)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     The Genome of the Natural Genetic Engineer Agrobacterium
            tumefaciens C58
  JOURNAL   Science 294 (5550), 2317-2323 (2001)
   PUBMED   11743193
REFERENCE   2  (bases 1 to 12952)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
            98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .12952
                     /organism="Agrobacterium tumefaciens str. C58 (U.
                     Washington)"
                     /strain="C58"
                     /db_xref="taxon:180835"
     gene            1022. .1510
                     /gene="Atu4791"
     CDS             1022. .1510
                     /gene="Atu4791"
                     /note="putative; ORF located using Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL45585.1"
                     /db_xref="GI:17743303"
                     /translation="MTRSNSGLYVGTTFADTGVAGIWGGLGNVYRRNRFTDIAYIGVE
                     YETWNHPGADYNGTVFDGNTVANVRYGFVDAYKLMWTHDGSFKAGPAERSRRINRVLY
                     DNHFTRGTPGPVQSAGFVTMHPQNSWVNVGSTWSGFAMGNGGPKGGAVKSSSNAARPT
                     PD"
     gene            complement(2052. .2348)
                     /gene="Atu4792"
     CDS             complement(2052. .2348)
                     /gene="Atu4792"
                     /note="putative; ORF located using Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL45586.1"
                     /db_xref="GI:17743304"
                     /translation="MESAIVPVGYHMVRQRAVIILNGERGLQPVHVTQDALNDIQSPP
                     QCDAQRLAQFVDVFATIAASKIEKGMFAFDGQIWITATDVQVWLNLAAENGPAK"
     gene            complement(2572. .3831)
                     /gene="bme11"
                     /note="Atu4793"
     CDS             complement(2572. .3831)
                     /gene="bme11"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="exopolysaccharide production protein"
                     /protein_id="AAL45587.1"
                     /db_xref="GI:17743305"
                     /translation="MRFLTRSRGSSLMIASTTFLSLAFYQIPAAADGTNYKVATGDVL
                     TISVYGDPGLTGLFPVSADGTIGYPLLGNVNVVDQTVNEIGERISRDLASHVANRSVA
                     VAVKEYAPIFVVGDVQKPGKYEFRPGMIVLELFALGGGLREPNAQRDTSGIQLIAAQQ
                     EYEDMSMQLLSQDIRRVRLEAELNDTEFEYRGDEIGLVRDRAALEKIVASETSLFRLR
                     LSAQQDEKTNLEVQRQNFIQEIDTLQKSNVMRGQQLELLDMDVNASKELVTRGAASES
                     ALRERKRELLSMNQQVLESTSFLARAQQNKSEVERRILELKSKRHNDAATELRETSLD
                     IMRLKKKLAFSLQSMAEIGSTARRVSSLEDMIQTKFFIVRQVAGTYRETAADEHTQVR
                     AGDVVRVRLSVSGRDPVSAAVVTGSVN"
     gene            complement(3835. .5997)
                     /gene="bme12"
                     /note="Atu4794"
     CDS             complement(3835. .5997)
                     /gene="bme12"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="polysaccharide biosynthesis protein"
                     /protein_id="AAL45588.1"
                     /db_xref="GI:17743306"
```

/translation="MIQSSSSPGEQEGISRFISILRERRRAVFICFLASFAIAAIGYA
RTPIRYKAEATLALDVRKLQALPTESVVSPLPQESPVLRTELDIIGSRSMAERVLVIL
QKDRLLEGRQTNSDRATNDVTTLGNDAETSKKVDNLMTNVRVMNDGRSYTIYISYYAG
DPALAAKVANAFGEAYIDYQIDLQTTATRRVSAWLGERLVSLRTALEESERQASQFRE
KSGLINTGGLPLQSQRLSSLNVELTTLQAKLAGSRARLATALALQKSGDGLGITEVLS
SPTIQLLRAEQARVIRAIAQIDESGAVMSPERPQLSSQLSSLNKQIDSEIAQIIAALR
NEIEVNEKQRQEVEASLNDVQRAISTTDMAMVQSGQLDREAAANRSIYETYLARYKQT
IEQDGIATAEARIISSAMPSRKPTSPDAGLWGLVALLLGCSTGLATAFILNFTDRSVG
SITSLQQKTGLNVIGRIPRISFGEFQNKPGKFSSGAADFYSAFADLQTHLQLAESKAK
VIAFTSSADGEGKTFVIANLARSLAMSGVKLLLIDANLRNPALGAEFDVGSTSHVAQA
VSHEVTLDDIIQRDQASGVDIIIAEKSDMPPNLVLGSKRFAGLIVEARQRYELILIEA
PSVDHDLDLLRIATLSDAVSFVVRQDVADGTGMKNAVIKLRVAGRPITGIILNGIRRP
RKRQPPVWPRIFRDWNPHFKLRKRATSLRPPAVVHNER"

gene complement(6052. .6663)
/gene="bmel3"
/note="Atu4795"

CDS complement(6052. .6663)
/gene="bmel3"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, LuxR family"
/protein_id="AAL45589.1"
/db_xref="GI:17743307"
/translation="MLERECLAHSLNYQYADRCILTFSNLGEWQQARAEIGMPAAVLV
SAWPGPSAETNRELAEIVAAVSPAAVVVLSDDQDINTILDTIALGVRGYISSSVSVDV
CVQAISLAIAGGRFIPASSVMNMKSLFGLTAHPAIPPMKFTSRQSAIAEALRCGKANK
DIALELNLCESTIKVHIRNIMKKLGATNRTEVACKIGDMAYER"

gene complement(6849. .7493)
/gene="Atu4796"

CDS complement(6849. .7493)
/gene="Atu4796"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45590.1"
/db_xref="GI:17743308"
/translation="MSLASAGDRLSILCLGAHSDDIEIGVGGTVLSLLSKGVLLDVYW
CVLSAGNERALEARSSAEAFLAKAASYSIELASFQDSYFPSQSREIKSWLLTLRDKVS
PDIIFTHSREDAHQDHREINRLTSNIFRNHLILEYEIPKWDGDFGRRNAYVPLSEDAM
NMKVELLLQHFGTQRSKDWFDADTFRGLARLRGMECRAPTRFAEAFTVRKLTFA"

gene complement(7499. .8311)
/gene="rfbF"
/note="Atu4797"

CDS complement(7499. .8311)
/gene="rfbF"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="glucose-1-phosphate cytidylyltransferase"
/protein_id="AAL45591.1"
/db_xref="GI:17743309"
/translation="MKVVLFCGGRGTRIREYSESVPKPLIPLGSQPIMRHVMQYYARY
GHEDFILCLGYKAHVIKEFFLNSRPESYADCVISSPGRVELLEELPRDWRVSLIDTGI
WRNIGERLWAVRHHLKGEKMFLANYSDGLTDVDLDDMIAQFEKSDKLACFLAVRPPLT
YHLAEIDESGKVREFRSSETSDMWINGGYFLMRPEIFDYMREGEELVLEPFSRLISED
KLMAYKHEGFWRSMDTLRDWQTLEAMVERGDMPWIGDDASMARKSSNAKLLP"

gene complement(8308. .9600)
/gene="Atu4798"

CDS complement(8308. .9600)
/gene="Atu4798"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45592.1"
/db_xref="GI:17743310"
/translation="MNNEAVFSARLSSTQHVGLCRLCNSPLVHTFVDLGMSPPCESFV
AADAANDVEPFYPLHAFVCDECFLVQLQEYVAPENIFTEYAYFSSFSDSWVAHAKRYC
DMVIERFALGQSSFVVELASNDGYLLQHFLTSNIPMLGIEPAVNVAKVAIGKGIPTLT
EFFNEALATDMAARGQKADLIIGNNVLAQVPDINDFVAGMKTLLKPEGVITLEFPHIE
KLIAENQFDTIYHEHFSYFSLLTIEKMARRHGLKVFDVEEIPTHGGSLRVFFSHEDIR
FPREARVDSLLARELNAGLDKIETYTAFAEAVRQTKRNLLSFLIRLKEMRKSICAYGA
PGKGNTLLNYCGIGTDFIDFAVDRNPYKHGRLTPGMHIPIRPVSEIPHMKPDYVLILP
WNLKNEIVAQMKDIRNWGGKFIVPIPDISIIDPKELVQ"

gene 9905. .10432
/gene="rfbC"
/note="Atu4799"

CDS 9905. .10432
/gene="rfbC"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="dTDP-4-dehydrorhamnose 3,5-epimerase"
/protein_id="AAL45593.1"
/db_xref="GI:17743311"
/translation="MLFEKTKIEGLLVVNTIPVRDSRGSFARTFCKREFESNALVSDF
SQHSLSLSKLKHTIRGLHFQRPPHQETKLVSCVKGSIWDVAVDLRRGSPTYLQWVGTV
LSADNGAQFYIPEGFAHGFQSLSDDVAVSYLISVPYAEASAAGIRHDEPAIGVEWPAP
ASVISEKDLGWPAWQ"

Query Match 70.0%; Score 16.8; DB 1; Length 12952;
Best Local Similarity 90.0%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 tagcgcccgtaacgatcggc 22
| ||||||| ||||||||||
Db 11095 TTGCGCCCGGAACGATCGGC 11076

Search completed: June 23, 2002, 06:31:50
Job time: 63842 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 14:59:54 ; Search time 652.13 Seconds
(without alignments)
63.187 Million cell updates/sec

Title: US-09-747-514A-2
Perfect score: 24
Sequence: 1 ggtagcgcccgtaacgatcggccg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	17.6	73.3	927	22	AAH68133	C glutamicum codin
	2	17.6	73.3	309400	22	AAH68534	C glutamicum codin
c	3	16.6	69.2	23128	23	AAS59552	Propionibacterium
	4	16.2	67.5	983	22	ABA77078	Proliferative glom
c	5	16.2	67.5	1386	22	AAH73679	Mouse nuclear prot
c	6	16.2	67.5	4383	23	ABL07815	Drosophila melanog
	7	16.2	67.5	5660	24	ABL34357	Human immune syste
	8	16.2	67.5	7690	23	ABL07814	Drosophila melanog
	9	16	66.7	671	21	AAF13425	Aspergillus oryzae
c	10	15.8	65.8	549	16	AAQ80347	DNA fragment funct
c	11	15.8	65.8	549	18	AAT92906	Coryneform 549 bp
c	12	15.8	65.8	549	19	AAV15033	Promoter from B.fl
c	13	15.8	65.8	551	16	AAQ80346	DNA fragment funct
c	14	15.8	65.8	551	18	AAT92905	Coryneform 551 bp
c	15	15.8	65.8	551	19	AAV15032	Promoter from B.fl
	16	15.8	65.8	825	22	ABA89279	Escherichia coli p
c	17	15.8	65.8	6921	22	ABA89273	Escherichia coli p
	18	15.6	65.0	750	20	AAV87760	EST clone EQ215.
c	19	15.6	65.0	1338	21	AAF12622	Aspergillus oryzae
	20	15.6	65.0	2475	23	AAS93040	DNA encoding novel
c	21	15.6	65.0	3042	23	AAS93042	DNA encoding novel
	22	15.6	65.0	3042	23	AAS94460	DNA encoding novel
	23	15.6	65.0	3266	23	AAS92874	DNA encoding novel
c	24	15.6	65.0	6235	24	ABL32787	Human immune syste
	25	15.6	65.0	6650	23	ABL03869	Drosophila melanog
	26	15.6	65.0	31562	23	ABL03868	Drosophila melanog
	27	15.6	65.0	4403765	22	AAI99683	Mycobacterium tube
c	28	15.6	65.0	4403765	22	AAI99683	Mycobacterium tube
	29	15.6	65.0	4411529	22	AAI99682	Mycobacterium tube
c	30	15.4	64.2	2055	23	AAS74960	DNA encoding novel
c	31	15.4	64.2	21429	22	AAF57362	Murine Cdc25A comp
c	32	15.2	63.3	845	17	AAT11323	DNA encoding recom
c	33	15.2	63.3	873	22	AAF26345	P. putida oxygenas
c	34	15.2	63.3	897	22	AAH65789	C glutamicum codin
	35	15.2	63.3	931	22	ABA77051	Proliferative glom
c	36	15.2	63.3	984	22	AAF71599	Corynebacterium gl
c	37	15.2	63.3	1154	21	AAC44106	Arabidopsis thalia
	38	15.2	63.3	1164	10	AAN92631	DNA fragment conta
	39	15.2	63.3	1229	23	AAS93349	DNA encoding novel
c	40	15.2	63.3	1314	21	AAA64885	Bordetella pertuss
	41	15.2	63.3	1630	15	AAQ66684	Acy B2 gene - enco
c	42	15.2	63.3	1632	22	AAF61090	P. putida KT2440-a
c	43	15.2	63.3	2033	21	AAC40796	Arabidopsis thalia
	44	15.2	63.3	2599	23	ABL05025	Drosophila melanog
	45	15.2	63.3	2729	10	AAN92629	DNA fragment conta

ALIGNMENTS

RESULT 1
AAH68133/c
ID AAH68133 standard; DNA; 927 BP.
XX
AC AAH68133;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 3168.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG92914.

XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 8; SEQ ID NO: 3168; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 927 BP; 187 A; 253 C; 277 G; 210 T; 0 other;

Query Match 73.3%; Score 17.6; DB 22; Length 927;
Best Local Similarity 83.3%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggtagcgcccgtaacgatcggccg 24
||| ||||||| ||| |||||| |
Db 687 GGTTGCGCCCGGAACAATCGGCTG 664

RESULT 2
AAH68534
ID AAH68534 standard; DNA; 309400 BP.
XX
AC AAH68534;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7069.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Disclosure; SEQ ID NO: 7069; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

Query Match 73.3%; Score 17.6; DB 22; Length 309400;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggtagcgcccgtaacgatcggccg 24
||| ||||||| ||| |||||| |
Db 70454 ggttgcgcccggaacaatcggctg 70477

RESULT 3
AAS59552/c
ID AAS59552 standard; DNA; 23128 BP.
XX
AC AAS59552;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein encoding DNA #47.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Claim 1; SEQ ID No 47; 1069pp; English.
XX
CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
CC in infections of bone, joints and the central nervous system, however it

CC is particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU51663-AAU51893 and AAU67535.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 23128 BP; 4349 A; 6746 C; 7113 G; 4908 T; 12 other;

Query Match 69.2%; Score 16.6; DB 23; Length 23128;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggtagcgcccgtaacgatcggcc 23
| | |||||||||||| ||||||
Db 19964 GCTCACGCCCGTAACGCTCGGCC 19942

RESULT 4
ABA77078
ID ABA77078 standard; DNA; 983 BP.
XX
AC ABA77078;
XX
DT 24-JAN-2002 (first entry)
XX
DE Proliferative glomerular nephritis-associated gene sequence SEQ ID:85.
XX
KW Rat; proliferative glomerular nephritis-associated gene; TRDH;
KW stromal cell derived factor-2; prostacyclin-stimulation factor;
KW TSC-22 like protein 2; kidney disease; diagnosis; kidney disorder;
KW proliferative glomerular nephritis; ds.
XX
OS Rattus norvegicus.
XX
PN WO200173022-A1.
XX
PD 04-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-JP02623.
XX
PR 29-MAR-2000; 2000JP-0090137.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Takeuchi K, Sekine S, Kikuchi Y, Sakurada K;
XX
DR WPI; 2001-616500/71.
XX
PT New DNA having increased expression in kidney tissues affected by
PT proliferative glomerular nephritis for diagnosis and treatment of
PT kidney disease and promotion of repair of damaged kidney tissue -
XX
PS Claim 16; Page 222-223; 314pp; Japanese.
XX
CC The present invention describes polynucleotide sequences of rat origin
CC which encode proteins having increased expression in kidney tissues
CC affected by proliferative glomerular nephritis. The proliferative
CC glomerular nephritis-associated polynucleotide and protein sequence have
CC nephrotropic activity. The polynucleotides can be used in the diagnosis,
CC treatment and prevention of kidney disease, especially of proliferative
CC glomerular nephritis, and in the repair of tissues damaged by kidney
CC disease. ABA77002 to ABA77154 and AAG68138 to AAG68147 represent
CC sequences given in the exemplification of the present invention.
XX
SQ Sequence 983 BP; 301 A; 205 C; 206 G; 250 T; 21 other;

Query Match 67.5%; Score 16.2; DB 22; Length 983;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gtagcgcccgtaacgatcggcc 23
| || ||||| | |||||||||
Db 805 gnagggcccgcaccgatcggcc 826

RESULT 5
AAH73679/c
ID AAH73679 standard; cDNA; 1386 BP.
XX
AC AAH73679;
XX
DT 17-OCT-2001 (first entry)
XX
DE Mouse nuclear protein 18A cDNA.
XX
KW Mouse; SP-10; promoter; antiinfertility; nuclear protein; sperm-specific;
KW intra-acrosomal protein; fertilisation; male fertility; contraceptive;
KW 18A; ss.
XX
OS Mus musculus.
XX
PN WO200153488-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01954.
XX
PR 19-JAN-2000; 2000US-0176886.
XX
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI Herr JC, Reddi PP, Acharya K;
XX
DR WPI; 2001-442257/47.
XX
PT New purified nuclear protein having a molecular weight of 45-50kDa and
PT a pI of 5.3 (+/- 0.2), for developing inhibitory agents for modulating
PT male fertility, which may function as male contraceptives -
XX
PS Claim 5; Page 23-24; 26pp; English.
XX
CC The invention relates to a purified nuclear protein having a molecular
CC weight of 45-50kDa and a pI of 5.3 (+/- 0.2), that specifically binds
CC to the regulatory sequences of the SP-10 gene. SP-10 is a
CC sperm-specific, intra-acrosomal protein that may have important function
CC in fertilisation. The protein and nucleic acids encoding it are used
CC to develop inhibitory agents for modulating male fertility which may
CC function as male contraceptives. The present sequence is a nucleic acid
CC from clone 18A of a mouse testis expression cDNA library. The clone was
CC isolated by screening the library using a radio-labelled probe
CC representing the core SP-10 promoter domain. The present sequence
CC encodes a protein that interacts with the SP-10 promoter.
XX
SQ Sequence 1386 BP; 365 A; 355 C; 360 G; 286 T; 20 other;

Query Match 67.5%; Score 16.2; DB 22; Length 1386;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggtagcgcccgtaacgatcggc 22
|| || ||||||| |||||||
Db 956 GGCAGGGCCCGTANAGATCGGC 935

```
RESULT   6
ABL07815/c
ID   ABL07815 standard; cDNA; 4383 BP.
XX
AC   ABL07815;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 17927.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB63712.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 17927; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4383 BP; 1261 A; 1159 C; 1192 G; 771 T; 0 other;


  Query Match             67.5%;  Score 16.2;  DB 23;  Length 4383;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcgg 21
           |||||||||||| ||| ||| |
Db    2292 GGTAGCGCCCGGAACTATCTG 2272


RESULT   7
ABL34357
ID   ABL34357 standard; DNA; 5660 BP.
XX
AC   ABL34357;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2330.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 2330; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5660 BP; 1177 A; 236 C; 1722 G; 2523 T; 2 other;


  Query Match             67.5%;  Score 16.2;  DB 24;  Length 5660;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 agcgcccgtaacgatcggccg 24
           ||||| ||||||| |||| ||
Db    4833 agcgcgcgtaacgttcgggcg 4853


RESULT   8
ABL07814
ID   ABL07814 standard; cDNA; 7690 BP.
XX
AC   ABL07814;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 17924.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
```

PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB63711.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 17924; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7690 BP; 1779 A; 1754 C; 1833 G; 2324 T; 0 other;

Query Match 67.5%; Score 16.2; DB 23; Length 7690;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggtagcgcccgtaacgatcgg 21
||||||||||| ||| ||| |
Db 3092 ggtagcgcccggaactatctg 3112

RESULT 9
AAF13425
ID AAF13425 standard; cDNA; 671 BP.
XX
AC AAF13425;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5948.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2456; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 671 BP; 183 A; 179 C; 175 G; 132 T; 2 other;

Query Match 66.7%; Score 16; DB 21; Length 671;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggtagcgcccgtaacgatcggccg 24
| |||||| ||||| | || ||||
Db 522 gctagcgcacgtaaagttcagccg 545

RESULT 10
AAQ80347/c
ID AAQ80347 standard; DNA; 549 BP.
XX
AC AAQ80347;
XX
DT 11-SEP-1995 (first entry)
XX
DE DNA fragment functional as a promoter in Coryneform bacteria.
XX
KW Promoter; Coryneform bacterium; Brevibacterium flavum; enzyme;
KW gene expression; biosynthesis; metabolism; bio-substance; vitamin;
KW amino acid; organic acid; lipid; protein; fat; oil; antibiotic; ds.
XX
OS Brevibacterium flavum MJ-233 (FERM BP-1497).
XX
PN EP629699-A.
XX
PD 21-DEC-1994.
XX
PF 07-JUN-1994; 94EP-0108738.
XX
PR 15-JUN-1993; 93US-0076091.
XX
PA (MITP) MITSUBISHI PETROCHEMICAL CO LTD.
XX
PI Yukawa H, Zupancic TJ;
XX
DR WPI; 1995-024246/04.
XX
PT DNA fragment from coryneform bacteria with promoter activity -

PT expresses the gene of interest in coryneform bacteria at a high
PT rate and has a greater promoter strength than a tac promoter
XX
PS Claim 13; Page 58; 67pp; English.
XX
CC The promoter DNA fragment can be used to express a gene of interest
CC in Coryneform bacteria at a high level. It may also be used for
CC controlling the level of expression of a gene of interest in
CC Coryneform bacteria. The promoter fragment may be used to express
CC gene products such as enzymes involved in biosynthesis and
CC metabolism of bio-substances e.g. amino acids, organic acids,
CC vitamins, lipids; and enzymes involved in the biosynthesis and
CC metabolism of bioactive substances such as proteins, fats and oils
CC and antibiotics. Twenty sequences with such promoter activity
CC are described (See also AAQ80333-52). This promoter is controllable
CC by replacing ethanol in the culture medium with glucose.
XX
SQ Sequence 549 BP; 122 A; 131 C; 143 G; 153 T; 0 other;

Query Match 65.8%; Score 15.8; DB 16; Length 549;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 agcgcccgtaacgatcggc 22
|||||| | ||||||||||
Db 177 AGCGCCTGGAACGATCGGC 159

RESULT 11
AAT92906/c
ID AAT92906 standard; DNA; 549 BP.
XX
AC AAT92906;
XX
DT 27-APR-1998 (first entry)
XX
DE Coryneform 549 bp promoter fragment.
XX
KW Promoter; coryneform bacterium; gene expression; ds.
XX
OS Brevibacterium flavum MJ-233 (FERM BP-1497).
XX
PN EP803575-A1.
XX
PD 29-OCT-1997.
XX
PF 07-JUN-1994; 94EP-0108382.
XX
PR 15-JUN-1993; 93US-0076091.
XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
PI Yukawa H, Zupancic TJ;
XX
DR WPI; 1997-515322/48.
XX
PT Promoters for gene expression in Coryneform bacteria - used for
PT controllable expression of genes by changing carbon source in
PT culture medium
XX
PS Claim 2; Page 48; 58pp; English.
XX
CC This novel isolated DNA fragment is functional as a promoter in
CC coryneform bacteria. It has greater strength in coryneform cells
CC than the tac promoter. The promoter function of the DNA fragment is
CC controllable by replacing ethanol with glucose in a culture medium
CC for a coryneform bacterium containing the DNA fragment. Claimed
CC promoter fragments (AAT92904-11) can be used for externally
CC controllable expression of genes in coryneform bacteria, e.g.
CC microbial, animal or plant genes encoding enzymes involved in the
CC biosynthesis or metabolism of substances such as amino acids,
CC organic acids, vitamins, lipids, proteins, fats, oils or
CC antibiotics. All are derived from Brevibacterium flavum MJ233
CC chromosomal DNA. They were obtained by inserting Alu-HaeII
CC restriction fragments of the DNA into the promoter probe shuttle
CC vector pPROBE17 at its EcoRV site.
XX
SQ Sequence 549 BP; 122 A; 131 C; 143 G; 153 T; 0 other;

Query Match 65.8%; Score 15.8; DB 18; Length 549;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 agcgcccgtaacgatcggc 22
|||||| | ||||||||||
Db 177 AGCGCCTGGAACGATCGGC 159

RESULT 12
AAV15033/c
ID AAV15033 standard; DNA; 549 BP.
XX
AC AAV15033;
XX
DT 28-MAY-1998 (first entry)
XX
DE Promoter from B.flavum MJ233 KG101.
XX
KW Promoter; coryneform bacteria; genetic engineering;
KW gene expression control; ds.
XX
OS Brevibacterium flavum.
XX
PN US5726299-A.
XX
PD 10-MAR-1998.
XX
PF 01-AUG-1994; 94US-0285641.
XX
PR 01-AUG-1994; 94US-0285641.
PR 03-JUN-1991; 91US-0709151.
PR 15-JUN-1993; 93US-0076091.
XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
PI Yukawa H, Zupancic TJ;
XX
DR WPI; 1998-192831/17.
XX
PT Coryneform bacteria promoter sequence(s) - controllable by change in
PT culture medium
XX
PS Claim 10; column 65-66; 40pp; English.
XX
CC This sequence represents a promoter from Brevibacterium flavum, and is an
CC example of the DNA sequence of the invention. The sequence is a DNA
CC fragment obtained from a coryneform bacteria, and which functions as a
CC promoter. The promoter's function is controlled by: (a) removing at least
CC one substance from a coryneform bacteria culture medium; (b) adding at
CC least one substance which is assimilable by coryneform bacteria; or
CC (c) both (a) and (b). The promoter is useful in genetic engineering for
CC controlling expression of genes of interest in coryneform bacteria.
XX
SQ Sequence 549 BP; 122 A; 131 C; 143 G; 153 T; 0 other;

Query Match 65.8%; Score 15.8; DB 19; Length 549;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 agcgcccgtaacgatcggc 22
|||||| | ||||||||||
Db 177 AGCGCCTGGAACGATCGGC 159

RESULT 13
AAQ80346/c
ID AAQ80346 standard; DNA; 551 BP.
XX
AC AAQ80346;
XX
DT 11-SEP-1995 (first entry)
XX
DE DNA fragment functional as a promoter in Coryneform bacteria.
XX
KW Promoter; Coryneform bacterium; Brevibacterium flavum; enzyme;
KW gene expression; biosynthesis; metabolism; bio-substance; vitamin;
KW amino acid; organic acid; lipid; protein; fat; oil; antibiotic; ds.
XX
OS Brevibacterium flavum MJ-233 (FERM BP-1497).
XX
PN EP629699-A.
XX
PD 21-DEC-1994.
XX
PF 07-JUN-1994; 94EP-0108738.
XX
PR 15-JUN-1993; 93US-0076091.
XX
PA (MITP) MITSUBISHI PETROCHEMICAL CO LTD.
XX
PI Yukawa H, Zupancic TJ;
XX
DR WPI; 1995-024246/04.
XX
PT DNA fragment from coryneform bacteria with promoter activity -
PT expresses the gene of interest in coryneform bacteria at a high
PT rate and has a greater promoter strength than a tac promoter
XX
PS Claim 12; Page 58; 67pp; English.
XX
CC The promoter DNA fragment can be used to express a gene of interest
CC in Coryneform bacteria at a high level. It may also be used for
CC controlling the level of expression of a gene of interest in
CC Coryneform bacteria. The promoter fragment may be used to express
CC gene products such as enzymes involved in biosynthesis and
CC metabolism of bio-substances e.g. amino acids, organic acids,
CC vitamins, lipids; and enzymes involved in the biosynthesis and
CC metabolism of bioactive substances such as proteins, fats and oils
CC and antibiotics. Twenty sequences with such promoter activity
CC are described (See also AAQ80333-52). This promoter is controllable
CC by replacing glucose in the culture medium with ethanol.
XX
SQ Sequence 551 BP; 126 A; 130 C; 138 G; 157 T; 0 other;

Query Match 65.8%; Score 15.8; DB 16; Length 551;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 agcgcccgtaacgatcggc 22
|||||| | ||||||||||
Db 176 AGCGCCTGGAACGATCGGC 158

RESULT 14
AAT92905/c
ID AAT92905 standard; DNA; 551 BP.
XX
AC AAT92905;
XX
DT 27-APR-1998 (first entry)
XX
DE Coryneform 551 bp promoter fragment.
XX
KW Promoter; coryneform bacterium; gene expression; ds.
XX
OS Brevibacterium flavum MJ-233 (FERM BP-1497).
XX
PN EP803575-A1.
XX
PD 29-OCT-1997.
XX
PF 07-JUN-1994; 94EP-0108382.
XX
PR 15-JUN-1993; 93US-0076091.
XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
PI Yukawa H, Zupancic TJ;
XX
DR WPI; 1997-515322/48.
XX
PT Promoters for gene expression in Coryneform bacteria - used for
PT controllable expression of genes by changing carbon source in
PT culture medium
XX
PS Claim 1; Page 48; 58pp; English.
XX
CC This novel isolated DNA fragment is functional as a promoter in
CC coryneform bacteria. It has greater strength in coryneform cells
CC than the tac promoter. The promoter function of the DNA fragment is
CC controllable by replacing glucose with ethanol in a culture medium
CC for a coryneform bacterium containing the DNA fragment. Claimed
CC promoter fragments (AAT92904-11) can be used for externally
CC controllable expression of genes in coryneform bacteria, e.g.
CC microbial, animal or plant genes encoding enzymes involved in the
CC biosynthesis or metabolism of substances such as amino acids,
CC organic acids, vitamins, lipids, proteins, fats, oils or
CC antibiotics. All are derived from Brevibacterium flavum MJ233
CC chromosomal DNA. They were obtained by inserting Alu-HaeII
CC restriction fragments of the DNA into the promoter probe shuttle
CC vector pPROBE17 at its EcoRV site.
XX
SQ Sequence 551 BP; 126 A; 130 C; 138 G; 157 T; 0 other;

Query Match 65.8%; Score 15.8; DB 18; Length 551;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 agcgcccgtaacgatcggc 22
|||||| | ||||||||||
Db 176 AGCGCCTGGAACGATCGGC 158

RESULT 15
AAV15032/c
ID AAV15032 standard; DNA; 551 BP.
XX
AC AAV15032;
XX
DT 28-MAY-1998 (first entry)
XX
DE Promoter from B.flavum MJ233 KE102.
XX
KW Promoter; coryneform bacteria; genetic engineering;
KW gene expression control; ds.
XX
OS Brevibacterium flavum.
XX
PN US5726299-A.
XX
PD 10-MAR-1998.
XX
PF 01-AUG-1994; 94US-0285641.
XX
PR 01-AUG-1994; 94US-0285641.
PR 03-JUN-1991; 91US-0709151.
PR 15-JUN-1993; 93US-0076091.

XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
PI Yukawa H, Zupancic TJ;
XX
DR WPI; 1998-192831/17.
XX
PT Coryneform bacteria promoter sequence(s) - controllable by change in
PT culture medium
XX
PS Claim 9; column 63-64; 40pp; English.
XX
CC This sequence represents a promoter from Brevibacterium flavum, and is an
CC example of the DNA sequence of the invention. The sequence is a DNA
CC fragment obtained from a coryneform bacteria, and which functions as a
CC promoter. The promoter's function is controlled by: (a) removing at least
CC one substance from a coryneform bacteria culture medium; (b) adding at
CC least one substance which is assimilable by coryneform bacteria; or
CC (c) both (a) and (b). The promoter is useful in genetic engineering for
CC controlling expression of genes of interest in coryneform bacteria.
XX
SQ Sequence 551 BP; 126 A; 130 C; 138 G; 157 T; 0 other;

Query Match 65.8%; Score 15.8; DB 19; Length 551;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 agcgcccgtaacgatcggc 22
|||||| | ||||||||||
Db 176 AGCGCCTGGAACGATCGGC 158

Search completed: June 23, 2002, 15:00:17
Job time: 81149 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 15:03:00 ; Search time 167.81 Seconds
(without alignments)
35.130 Million cell updates/sec

Title: US-09-747-514A-2
Perfect score: 24
Sequence: 1 ggtagcgcccgtaacgatcggccg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	15.8	65.8	549	1	US-08-076-091C-15	Sequence 15, Appl
c	2	15.8	65.8	549	1	US-08-285-641-15	Sequence 15, Appl
c	3	15.8	65.8	551	1	US-08-076-091C-14	Sequence 14, Appl
c	4	15.8	65.8	551	1	US-08-285-641-14	Sequence 14, Appl
	5	15.6	65.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	6	15.6	65.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	7	15.6	65.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	8	15.2	63.3	777	4	US-08-998-416-400	Sequence 400, App
c	9	15.2	63.3	845	1	US-08-266-451B-1	Sequence 1, Appli
c	10	15.2	63.3	845	2	US-08-748-725-1	Sequence 1, Appli
	11	15.2	63.3	1164	6	5240849-4	Patent No. 5240849
	12	15.2	63.3	2749	6	5240849-1	Patent No. 5240849
	13	15.2	63.3	16885	1	US-08-390-878-16	Sequence 16, Appl
c	14	15	62.5	1519	1	US-08-225-477B-2	Sequence 2, Appli
c	15	15	62.5	1519	5	PCT-US95-04353-2	Sequence 2, Appli
c	16	15	62.5	2325	1	US-08-314-309A-20	Sequence 20, Appl
	17	15	62.5	2556	4	US-08-976-259-61	Sequence 61, Appl
c	18	15	62.5	20235	1	US-07-642-734C-3	Sequence 3, Appli
c	19	15	62.5	20235	3	US-08-439-009A-3	Sequence 3, Appli
c	20	15	62.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	21	14.8	61.7	771	5	PCT-US95-12987-1	Sequence 1, Appli
c	22	14.8	61.7	771	5	PCT-US95-12987-3	Sequence 3, Appli
c	23	14.8	61.7	771	5	PCT-US95-12987-5	Sequence 5, Appli
c	24	14.8	61.7	1540	1	US-08-463-115-2	Sequence 2, Appli
c	25	14.8	61.7	1540	1	US-08-465-388-2	Sequence 2, Appli
	26	14.8	61.7	2859	2	US-08-637-763B-7	Sequence 7, Appli
	27	14.8	61.7	2859	3	US-09-170-354-7	Sequence 7, Appli
c	28	14.8	61.7	3160	4	US-08-936-165A-255	Sequence 255, App
	29	14.6	60.8	1171	4	US-09-056-556-164	Sequence 164, App
c	30	14.6	60.8	1400	1	US-07-989-845-1	Sequence 1, Appli
c	31	14.6	60.8	1400	5	PCT-US93-11298-1	Sequence 1, Appli
c	32	14.6	60.8	1749	4	US-09-516-914-22	Sequence 22, Appl
c	33	14.6	60.8	1923	4	US-09-294-841-1	Sequence 1, Appli
	34	14.6	60.8	2065	4	US-09-319-989-5	Sequence 5, Appli
	35	14.6	60.8	2394	4	US-09-319-989-7	Sequence 7, Appli
c	36	14.6	60.8	2754	2	US-09-028-361A-1	Sequence 1, Appli
	37	14.6	60.8	3081	4	US-09-319-989-9	Sequence 9, Appli
c	38	14.6	60.8	3186	1	US-08-701-846-1	Sequence 1, Appli
	39	14.6	60.8	3360	4	US-09-319-989-8	Sequence 8, Appli
c	40	14.6	60.8	3572	2	US-08-713-815A-2	Sequence 2, Appli
c	41	14.6	60.8	3807	2	US-08-816-755-1	Sequence 1, Appli
c	42	14.6	60.8	3807	4	US-09-090-673-1	Sequence 1, Appli
c	43	14.6	60.8	8878	1	US-08-759-444-2	Sequence 2, Appli
c	44	14.6	60.8	9880	3	US-08-680-897-1	Sequence 1, Appli
c	45	14.6	60.8	14272	4	US-09-516-914-23	Sequence 23, Appl

ALIGNMENTS

```
RESULT 1
US-08-076-091C-15/c
; Sequence 15, Application US/08076091C
; Patent No. 5693781
; GENERAL INFORMATION:
;   APPLICANT: Zupancic, Thomas J.
;   APPLICANT: Yukawa, Hideaki
;   TITLE OF INVENTION: PROMOTER DNA FRAGMENT FROM CORNYNEFORM
;   TITLE OF INVENTION: BACTERIA
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
;     STREET: 2100 Pennsylvania Avenue, N.W.
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: United States
;     ZIP: 20037-3202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/076,091C
;     FILING DATE: 15-JUN-1993
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/709,151
;     FILING DATE: 29-MAY-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Nakamura, Dean H.
;     REGISTRATION NUMBER: 33,981
;     REFERENCE/DOCKET NUMBER: Q-32690
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202) 293-7060
;     TELEFAX: (202) 293-7860
;     TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 15:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 549 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: genomic DNA
;   ORIGINAL SOURCE:
;     ORGANISM: Brevibacterium flavum
;     STRAIN: MJ233
;   FEATURE:
;     NAME/KEY: promoter
;     LOCATION: 1-549
```

```
;      IDENTIFICATION METHOD:  experiment
US-08-076-091C-15


  Query Match             65.8%;  Score 15.8;  DB 1;  Length 549;
  Best Local Similarity   89.5%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 agcgcccgtaacgatcggc 22
           |||||| | ||||||||||
Db     177 AGCGCCTGGAACGATCGGC 159


RESULT   2
US-08-285-641-15/c
; Sequence 15, Application US/08285641
; Patent No. 5726299
;  GENERAL INFORMATION:
;    APPLICANT:  Zupancic, Thomas J.
;    APPLICANT:  Yukawa, Hideaki
;    TITLE OF INVENTION:  PROMOTER DNA FRAGMENT FROM CORNYNEFORM BACTERIA
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:
;      COMPUTER:
;      OPERATING SYSTEM:
;      SOFTWARE:
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/285,641
;      FILING DATE:  1-AUGUST-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076,091
;      FILING DATE:  15-JUN-1993
;      APPLICATION NUMBER:  US 07/709,151
;      FILING DATE:  29-MAY-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  549  base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Brevibacterium flavum
;      STRAIN:  MJ-233
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1-549
;      IDENTIFICATION METHOD:  experiment
US-08-285-641-15


  Query Match             65.8%;  Score 15.8;  DB 1;  Length 549;
  Best Local Similarity   89.5%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 agcgcccgtaacgatcggc 22
           |||||| | ||||||||||
Db     177 AGCGCCTGGAACGATCGGC 159
```

```
RESULT   3
US-08-076-091C-14/c
; Sequence 14, Application US/08076091C
; Patent No. 5693781
;  GENERAL INFORMATION:
;    APPLICANT:  Zupancic, Thomas J.
;    APPLICANT:  Yukawa, Hideaki
;    TITLE OF INVENTION:  PROMOTER DNA FRAGMENT FROM CORNYNEFORM
;    TITLE OF INVENTION:  BACTERIA
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,091C
;      FILING DATE:  15-JUN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/709,151
;      FILING DATE:  29-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakamura, Dean H.
;      REGISTRATION NUMBER:  33,981
;      REFERENCE/DOCKET NUMBER:  Q-32690
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  551  base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Brevibacterium flavum
;      STRAIN:  MJ233
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1-551
;      IDENTIFICATION METHOD:  experiment
US-08-076-091C-14


  Query Match             65.8%;  Score 15.8;  DB 1;  Length 551;
  Best Local Similarity   89.5%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 agcgcccgtaacgatcggc 22
           |||||| | ||||||||||
Db     176 AGCGCCTGGAACGATCGGC 158


RESULT   4
US-08-285-641-14/c
; Sequence 14, Application US/08285641
; Patent No. 5726299
;  GENERAL INFORMATION:
;    APPLICANT:  Zupancic, Thomas J.
;    APPLICANT:  Yukawa, Hideaki
;    TITLE OF INVENTION:  PROMOTER DNA FRAGMENT FROM CORNYNEFORM BACTERIA
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;       STREET:  2100 Pennsylvania Avenue, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  United States
;       ZIP:  20037-3202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:
;       COMPUTER:
;       OPERATING SYSTEM:
;       SOFTWARE:
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/285,641
;       FILING DATE:  1-AUGUST-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076,091
;       FILING DATE:  15-JUN-1993
;       APPLICATION NUMBER:  US 07/709,151
;       FILING DATE:  29-MAY-1991
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:
;       TELEFAX:
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  551  base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Brevibacterium flavum
;       STRAIN:  MJ-233
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  1-551
;       IDENTIFICATION METHOD:  experiment
US-08-285-641-14


  Query Match             65.8%;  Score 15.8;  DB 1;  Length 551;
  Best Local Similarity   89.5%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 agcgcccgtaacgatcggc 22
           |||||| | ||||||||||
Db     176 AGCGCCTGGAACGATCGGC 158


RESULT   5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             65.0%;  Score 15.6;  DB 4;  Length 4403765;
  Best Local Similarity   79.2%;  Pred. No. 39;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           | | ||  |||||||||| ||||||
Db 4027823 gctcgcttccgtaacgaccggccg 4027846


RESULT   6
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             65.0%;  Score 15.6;  DB 4;  Length 4403765;
  Best Local Similarity   85.7%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 tagcgcccgtaacgatcggcc 23
           |||| ||||| ||||||| ||
Db 1537086 TAGCACCCGTGACGATCGACC 1537066


RESULT   7
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
  Query Match             65.0%;  Score 15.6;  DB 4;  Length 4411529;
  Best Local Similarity   79.2%;  Pred. No. 39;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggccg 24
           | | ||  ||||||||| ||||||
Db 4035742 gctcgcttccgtaacgaccggccg 4035765


RESULT   8
US-08-998-416-400/c
; Sequence 400, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  400:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1300RP
US-08-998-416-400


  Query Match             63.3%;  Score 15.2;  DB 4;  Length 777;
  Best Local Similarity   85.0%;  Pred. No. 89;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 gcgcccgtaacgatcggccg 24
           ||||| ||| ||||||||| |
Db      87 GCGCCTGTAGCGATCGGCGG 68


RESULT   9
US-08-266-451B-1/c
; Sequence 1, Application US/08266451B
; Patent No. 5623054
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang et al.
;    TITLE OF INVENTION:  CRUCIFER AFT PROTEINS AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/266,451B
;      FILING DATE:  23-June-1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/219001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  845
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-266-451B-1


  Query Match             63.3%;  Score 15.2;  DB 1;  Length 845;
  Best Local Similarity   85.0%;  Pred. No. 89;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gtagcgcccgtaacgatcgg 21
           |||||||| ||||||| | |
Db     147 GTAGCGCCTGTAACGAGCTG 128


RESULT  10
US-08-748-725-1/c
; Sequence 1, Application US/08748725
; Patent No. 5859346
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang et al.
;    TITLE OF INVENTION:  CRUCIFER AFT PROTEINS AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/748,725
;      FILING DATE:
```

```
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/266,451
;      FILING DATE:  23-June-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/219002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  845
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-748-725-1


  Query Match             63.3%;  Score 15.2;  DB 2;  Length 845;
  Best Local Similarity   85.0%;  Pred. No. 89;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gtagcgcccgtaacgatcgg 21
          ||||||| ||||||| | |
Db    147 GTAGCGCCTGTAACGAGCTG 128


RESULT  11
5240849-4
;Patent No. 5240849
;    APPLICANT: Arisawa, Akira;Kawamura, Naoto;Kojima,Ikou;Okumura,
;Yasushi;Kazuhiko, Okamura;Hiroshi, Tone;Okamura Rokuro
;    TITLE OF INVENTION: DNA CODING FOR ENZYME CAPABLE OF ACLYATING
;THE "4"-POSITION OF MACROLIDE ANTIBIOTIC
;    NUMBER OF SEQUENCES: 5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/06/353,323
;      FILING DATE: 23-MAY-1989
;SEQ ID NO:4:
;      LENGTH: 1164
5240849-4


  Query Match             63.3%;  Score 15.2;  DB 6;  Length 1164;
  Best Local Similarity   85.0%;  Pred. No. 89;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 gcgcccgtaacgatcggccg 24
          ||| |||| ||||||||| |
Db    506 gcgaccgtcacgatcggcag 525


RESULT  12
5240849-1
;Patent No. 5240849
;    APPLICANT: Arisawa, Akira;Kawamura, Naoto;Kojima,Ikou;Okumura,
;Yasushi;Kazuhiko, Okamura;Hiroshi, Tone;Okamura Rokuro
;    TITLE OF INVENTION: DNA CODING FOR ENZYME CAPABLE OF ACLYATING
;THE "4"-POSITION OF MACROLIDE ANTIBIOTIC
;    NUMBER OF SEQUENCES: 5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/06/353,323
;      FILING DATE: 23-MAY-1989
;SEQ ID NO:1:
;      LENGTH: 2749
5240849-1


  Query Match             63.3%;  Score 15.2;  DB 6;  Length 2749;
  Best Local Similarity   85.0%;  Pred. No. 90;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 gcgcccgtaacgatcggccg 24
          ||| |||| ||||||||| |
Db   2074 gcgaccgtcacgatcggcag 2093


RESULT  13
US-08-390-878-16
; Sequence 16, Application US/08390878
; Patent No. 5700683
;  GENERAL INFORMATION:
;    APPLICANT:  Stover, Charles K.
;    APPLICANT:  Mahairas, Gregory G.
;    TITLE OF INVENTION:  VIRULENCE-ATTENUATING GENETIC DELETIONS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Street Tower, 20th
;      STREET:  Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/390,878
;      FILING DATE:  17-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunter, Tom
;      REGISTRATION NUMBER:  38,498
;      REFERENCE/DOCKET NUMBER:  15371A-17
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/543/9600
;      TELEFAX:  415/543/5043
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16885 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-390-878-16


  Query Match             63.3%;  Score 15.2;  DB 1;  Length 16885;
  Best Local Similarity   85.0%;  Pred. No. 93;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 gcgcccgtaacgatcggccg 24
          ||||||| | |||||||| |
Db  16069 GCGCCCGGATCGATCGGCGG 16088


RESULT  14
US-08-225-477B-2/c
; Sequence 2, Application US/08225477B
; Patent No. 5635370
;  GENERAL INFORMATION:
;    APPLICANT:  Susan Hockfield
;    APPLICANT:  Diane M. Jaworski
;    TITLE OF INVENTION:  BEHAB, A Brain Hya-
;    TITLE OF INVENTION:  luronan-Binding Protein
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  St. Onge Steward Johnston & Reens
;      STREET:  986 Bedford Street
;      CITY:  Stamford
```

```
;      STATE:  CT
;      COUNTRY:  United States
;      ZIP:  06905
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" 1.44 Mb diskette
;      COMPUTER:  IBM PC
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  Word Processor
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/225,477B
;      FILING DATE:  April 8, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary M. Krinsky
;      REGISTRATION NUMBER:  32423
;      REFERENCE/DOCKET NUMBER:  1751-P0004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  203-324-6155
;      TELEFAX:  203-327-1096
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1519 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  DNA encoding a protein
;    FRAGMENT TYPE:  entire sequence
;    IMMEDIATE SOURCE:  cat cortex
;    FEATURE:
;      NAME/KEY:  cat brain BEHAB
US-08-225-477B-2


  Query Match             62.5%;  Score 15;  DB 1;  Length 1519;
  Best Local Similarity   78.3%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggcc 23
           |||||||| ||   | |||||||
Db    1209 GGTAGCGCACGCTGCCATCGGCC 1187


RESULT  15
PCT-US95-04353-2/c
; Sequence 2, Application PC/TUS9504353
;  GENERAL INFORMATION:
;    APPLICANT:  Susan Hockfield
;    APPLICANT:  Diane M. Jaworski
;    TITLE OF INVENTION:  BEHAB, A Brain Hya-
;    TITLE OF INVENTION:  luronan-Binding Protein
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  St. Onge Steward Johnston & Reens
;      STREET:  986 Bedford Street
;      CITY:  Stamford
;      STATE:  CT
;      COUNTRY:  United States
;      ZIP:  06905
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" 1.44 Mb diskette
;      COMPUTER:  IBM PC
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  Word Processor
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04353
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/225,477
;      FILING DATE:  April 8, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary M. Krinsky
;      REGISTRATION NUMBER:  32423
;      REFERENCE/DOCKET NUMBER:  1751-P0004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  203-324-6155
;      TELEFAX:  203-327-1096
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1519 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  DNA encoding a protein
;    FRAGMENT TYPE:  entire sequence
;    IMMEDIATE SOURCE:  cat cortex
;    FEATURE:
;      NAME/KEY:  cat brain BEHAB
PCT-US95-04353-2


  Query Match             62.5%;  Score 15;  DB 5;  Length 1519;
  Best Local Similarity   78.3%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 ggtagcgcccgtaacgatcggcc 23
           |||||||| ||   | |||||||
Db    1209 GGTAGCGCACGCTGCCATCGGCC 1187


Search completed: June 23, 2002, 15:09:02
Job time: 78566 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 06:31:50 ; Search time 2161.72 Seconds
(without alignments)
232.332 Million cell updates/sec

Title: US-09-747-514A-3
Perfect score: 24
Sequence: 1 ccgttcgggcgcgaggaagcccgg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	24	100.0	3123	1	PAU38241	U38241 Pseudomonas
c 2	24	100.0	13075	1	AE004946	AE004946 Pseudomon
c 3	20.8	86.7	110000	2	LMFLCHR16_03	Continuation (4 of
c 4	19.4	80.8	315000	1	RME603644	AL603644 Rhizobium
c 5	19.2	80.0	9000	1	SC9A10	AL031260 Streptomy
6	18.8	78.3	36874	9	AC005336	AC005336 Homo sapi
c 7	18.8	78.3	39877	9	AC020950	AC020950 Homo sapi
8	18.8	78.3	123693	9	AL162853	AL162853 Human DNA
9	18.8	78.3	151764	9	AC016396	AC016396 Homo sapi
10	18.8	78.3	182433	2	AL358155	AL358155 Homo sapi
c 11	18.2	75.8	960	33	AC077474	Ac077474 Giardia i
c 12	18.2	75.8	2214	5	AF043641	AF043641 Barbus ba
13	18.2	75.8	21057	3	LMFL673	AL135898 Leishmani
14	18.2	75.8	33311	9	AC004235	AC004235 Homo sapi
15	18.2	75.8	43903	2	AC004396	AC004396 Pseudomon
c 16	18.2	75.8	63734	1	AF127374	AF127374 Streptomy
17	18.2	75.8	121245	9	AC034240	AC034240 Homo sapi
18	18.2	75.8	163731	9	AC008949	AC008949 Homo sapi
c 19	17.8	74.2	494	11	G15996	G15996 human STS C
c 20	17.8	74.2	110000	2	LMFLCHR36_33	Continuation (34 o
21	17.8	74.2	132888	2	AP004652	AP004652 Oryza sat
22	17.8	74.2	135373	2	AP004695	AP004695 Oryza sat
23	17.6	73.3	771	33	AC039166	Ac039166 Giardia i
24	17.6	73.3	924	33	AC040499	Ac040499 Giardia i
c 25	17.6	73.3	961	33	AC089938	Ac089938 Giardia i
26	17.6	73.3	1449	1	CP16SRNB	X73444 C.puniceum
c 27	17.6	73.3	2672	1	AF159692	AF159692 Myxococcu
28	17.6	73.3	4656	1	ECCBRABCD	X87208 E.chrysanth
c 29	17.6	73.3	5341	9	HSSMAD2S1	U78726 Homo sapien
30	17.6	73.3	5439	1	AF377338	AF377338 Myxococcu
31	17.6	73.3	5869	1	AF111939	AF111939 Sinorhizo
c 32	17.6	73.3	13409	1	AE004466	AE004466 Pseudomon
33	17.6	73.3	13570	2	AC109008	AC109008 Rattus no
34	17.6	73.3	71864	2	AC068962	AC068962 Homo sapi
c 35	17.6	73.3	110000	2	AL607039_0	AL607039 Mus muscu
36	17.6	73.3	117004	2	AC108090	AC108090 Homo sapi
37	17.6	73.3	131613	2	AC026793	AC026793 Homo sapi
38	17.6	73.3	141706	9	AC010491	AC010491 Homo sapi
c 39	17.6	73.3	141894	9	AC093211	AC093211 Homo sapi
40	17.6	73.3	162839	2	AC106368	AC106368 Rattus no
c 41	17.6	73.3	164653	2	AC094444	AC094444 Rattus no
c 42	17.6	73.3	184777	9	AC026749	AC026749 Homo sapi
43	17.6	73.3	190048	2	AC094926	AC094926 Rattus no
44	17.6	73.3	340900	1	SME591791	AL591791 Sinorhizo
45	17.6	73.3	344050	1	MLEPRTN1	AL583917 Mycobacte

ALIGNMENTS

RESULT 1
PAU38241/c
LOCUS PAU38241 3123 bp DNA linear BCT 05-OCT-1996
DEFINITION Pseudomonas aeruginosa orotate phophoribosyl transferase (pyrE), catabolite repression control protein (crc) and RNasePH (rph) genes, complete cds.
ACCESSION U38241 L12038
VERSION U38241.1 GI:1079660
KEYWORDS .
SOURCE Pseudomonas aeruginosa strain=PAO1.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 3123)
AUTHORS MacGregor,C.H., Arora,S.K., Hager,P.W., Dail,M.B. and Phibbs,P.V. Jr.
TITLE The nucleotide sequence of the Pseudomonas aeruginosa pyrE-crc-rph region and the purification of the crc gene product
JOURNAL J. Bacteriol. 178 (19), 5627-5635 (1996)
MEDLINE 96421989
REFERENCE 2 (bases 1 to 3123)
AUTHORS Hager,P.W. and Phibbs,P.V. Jr.
TITLE Direct Submission

```
  JOURNAL   Submitted (11-OCT-1995) Paul W. Hager, Microbiology & Immunology,
            East Carolina University, Greenville, NC 27858, USA
COMMENT     On Nov 29, 1995 this sequence version replaced gi:496210.
FEATURES             Location/Qualifiers
     source          1. .3123
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
                     /map="11 min"
     gene            complement(414. .1055)
                     /gene="pyrE"
     CDS             complement(414. .1055)
                     /gene="pyrE"
                     /EC_number="2.4.2.10"
                     /function="catalytic activity: orotidine-5'-phosphate +
                     pyrophosphate = orotate + 5-phospho-alpha-d-ribose
                     1-diphosphate; Method: conceptual translation supplied by
                     author."
                     /codon_start=1
                     /transl_table=11
                     /product="orotate phosphoribosyl transferase"
                     /protein_id="AAC44427.1"
                     /db_xref="GI:1079661"
                     /translation="MQAYQRDFIRFAIERGVLRFGEFTLKSGRTSPYFFNAGLFDSGL
                     ALARLGRFYAEAVIDSGIDFDVLFGPAYKGIPLAATTAVALAEQHQRDLPWCFNRKEA
                     KEHGEGGTLVGAPLSGRVLIIDDVITAGTAIREVMQIIDAQGARAAGVLIALNRQERG
                     KGELSAIQEVERDFGMPVVSIVSLEQVLEYLAEDAELKKHLPAVEAYRAQYGI"
     gene            1136. .1915
                     /gene="crc"
     CDS             1136. .1915
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAC44428.1"
                     /db_xref="GI:1079662"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(2394. .3113)
                     /gene="rph"
     CDS             complement(2394. .3113)
                     /gene="rph"
                     /note="similar to ribonuclease PH"
                     /codon_start=1
                     /transl_table=11
                     /product="RNase PH"
                     /protein_id="AAC44429.1"
                     /db_xref="GI:1079663"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
                     ESGVPRFLKGQGQGWLTAEYGMLRRSTGERNQREASRGKQGGRTLEIQRLIDRSLRAA
                     LDLSKLGENTLYIDCDVIQTDGGTRTASITGATVSLIDALAVLKKRVALKGNPLKQIV
                     AAVSVGIYQGVPVLDLDYLEDSAAESDLNVVMTDCRRFIEVQGTAEGAPFRPAELNAM
                     LELAQQGMQELVRTQRAALAE"
BASE COUNT      551 a   1020 c    989 g    563 t
ORIGIN


  Query Match           100.0%;  Score 24;  DB 1;  Length 3123;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcccgg 24
          ||||||||||||||||||||||||
Db   1660 CCGTTCGGGCGCGAGGAAGCCCGG 1637

RESULT   2
AE004946/c
LOCUS       AE004946               13075 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 507 of 529 of the complete
            genome.
ACCESSION   AE004946 AE004091
VERSION     AE004946.1  GI:9951650
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .13075
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            71. .850
                     /gene="crc"
                     /note="PA5332"
     CDS             71. .850
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAG08717.1"
                     /db_xref="GI:9951651"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(909. .1280)
                     /gene="PA5333"
     CDS             complement(909. .1280)
                     /gene="PA5333"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08718.1"
                     /db_xref="GI:9951652"
                     /translation="MNDVSEEVAATPGQEARQWAMFCHFAAFIGLVFPFGSLLGPLIV
                     WQLKRESDPFVDAQGKEALNFQITVALAMFVSFLLMLVVIGFFLLGLVCLAALVLTII
                     AGIKANEGKAYRYPFCLRLIK"
     gene            complement(1327. .2046)
                     /gene="rph"
                     /note="PA5334"
     CDS             complement(1327. .2046)
                     /gene="rph"
                     /codon_start=1
                     /transl_table=11
                     /product="ribonuclease PH"
                     /protein_id="AAG08719.1"
                     /db_xref="GI:9951653"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
```

COMP

ESGVPRFLKGQGQGWLTAEYGMLPRSTGERNQREASRGKQGGRTLEIQRLIGRSLRAA
LDLSKLGENTLYIDCDVIQADGGTRTASITGATVALIDALAVLKKRGALKGNPLKQMV
AAVSVGIYQGVPVLDLDYLEDSAAETDLNVVMTDAGGFIEVQGTAEGAPFRPAELNAM
LELAQQGMQELFELQRAALAE"

gene 2227. .3090
/gene="PA5335"

CDS 2227. .3090
/gene="PA5335"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08720.1"
/db_xref="GI:9951654"
/translation="MVHSMTAFARVEQAGTHGTLSWELRSVNHRYLEPHLRLPDAFRD
LEGAVREALRQGLSRGKVECTLRFAEETAGKSLQVDQERARQLVAAAEGVAALIRQPA
PLDPLAVLAWPGVLVADSADPQALNAAALEAFGQALEQLKAGRSREGLELAKLLNDRL
DAMLVEVGNLRELVPTMLANQRQKILDRFAELKAELDPQRLEQELVLLAQKSDVAEEL
DRLATHVGEVRRVLKAGGAAGRRLDFLMQELNREANTLGSKAFDPRSTQAAVNLKVLI
EQMREQVQNIE"

gene 3149. .3760
/gene="gmk"
/note="PA5336"

CDS 3149. .3760
/gene="gmk"
/codon_start=1
/transl_table=11
/product="guanylate kinase"
/protein_id="AAG08721.1"
/db_xref="GI:9951655"
/translation="MSGTLYIVSAPSGAGKTSLVKALLDAAPEVRVSVSHTTRGMRPG
EVDGVNYHFTSREEFLAMLERNEFLEHAEVFGNLYGTSQRWVEKTLAEGLDLILEIDW
QGAQQVRRLMPEAQSIFILPPSQEALRQRLTNRGQDSDEVIERRMREAVSEMSHYVEY
DHLVINDDFAHALDDLKAIFRARQLRQDAQQQRHAELLGRLLA"

gene 3849. .4115
/gene="rpoZ"
/note="PA5337"

CDS 3849. .4115
/gene="rpoZ"
/codon_start=1
/transl_table=11
/product="RNA polymerase omega subunit"
/protein_id="AAG08722.1"
/db_xref="GI:9951656"
/translation="MARVTVEDCLDNVDNRFELVMLATKRARQLATGGKEPKVAWEND
KPTVVALREIASGLVDENVVQQEDIVEDEPLFAAFDDEANTEAL"

gene 4181. .6286
/gene="spoT"
/note="PA5338"

CDS 4181. .6286
/gene="spoT"
/codon_start=1
/transl_table=11
/product="guanosine-3',5'-bis(diphosphate)
3'-pyrophosphohydrolase"
/protein_id="AAG08723.1"
/db_xref="GI:9951657"
/translation="MPGIDAFADRLSSYLGPDQVNLVRRAYYYAEQAHDGQRRRSGEP
YVTHPLAVSNILADMHMDHQSLMAAMLHDVIEDTGIAKEALTAQFGETVSELVDGVSK
LTQMNFETKAEAQAENFQKMAMAMARDIRVILVKLADRLHNMRTLEVLSGEKRRRIAK
ETLEIYAPIANRLGMHTMRVEFEDLGFKAMHPMRAERIRQAVKKARGNRREIVGKIQE
SLVNCLAREGMEGQVIGREKHLYSIYQKMRGKRKAFNEIMDVYAFRIIVDKVDTCYRV
LGAVHSLYKPFPGRFKDYIAIPKANGYQSLHTTLFGMHGVPIEIQIRTREMEEMANHG
IAAHWLYKSNDETPKGTHARARQWVKGVLELQQRAGNSLEFIENVKIDLFPDEVYVFT
PKGRIMELPKGSTAVDFAYAVHTDVGNSCIACRINRRLAPLSEPLQSGQTVEIVTAPG
ARPNPAWLSFVVTGKARTHIRHALKLQRRSESINLGERLLNKTLTGFDSHLEKIPQER
IQAVLAEYRMEVFEDLLEEIGLGNRMAYVVARRLLSSEGEEAPSAEGPLAIRGTEGLV
LSYAKCCTPIPGDPIVGHLSAGKGMVVHLESCKNISEVRHNPEKCIQLSWAKDVTGEF
NVELRVELEHQRGLIALLATSVNAADGNIEKISMDERDGRISVVQLVVSVHDRVHLAR
VIKKLRALKGVIRITRLRS"

gene 6348. .6728
/gene="PA5339"

CDS 6348. .6728
/gene="PA5339"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08724.1"
/db_xref="GI:9951658"
/translation="MTKTVIHTDNAPAAIGTYSQAIKAGNTVYVSGQIPLDPKTMELV
EGFEEQTVQVFENLKAVIEASGGYLHDVAKLNIFLTDLSHFAKVNEIMGRYFTQPYPA
RAAIGVAALPKGAQVEMDAIVVLE"

gene 6783. .7514
/gene="PA5340"

CDS 6783. .7514
/gene="PA5340"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08725.1"
/db_xref="GI:9951659"
/translation="MRRVIFLAAAATLLAGCAGTADPSGTWINQAAIDAASKDGKLRE
ALLAYGPNLEWKLDSKAGEATFSNGFELGEGTLSKSDDEHWKVAFYGDDNQESLELDG
KELIQQASANGPEQRFRRLDPQPAASSPAGSGFERALYGSYLKGSWKIREGQGQGGKV
EFQANGLVSGLPGAERYALCLAGDCAAMSGDNDSIWLQQGNRGRELLFSLDDDELQLF
EAVNTAGANEMPSYVPGKRVWLLER"

gene complement(7521. .8141)
/gene="PA5341"

CDS complement(7521. .8141)
/gene="PA5341"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08726.1"
/db_xref="GI:9951660"
/translation="MFAAFVLVASTHFAALLSPGPDFFLLVRAALLRGRRQADGVAAG
IALANLLSMLLVLGLLASVPDSAHSALRWLQALGGLYFLWLAGQALLAQRRLEMPAQR
DVPSRGGYLRGLRDGLLASSLNPKLPIFYAGLFGVLARFSLPGWALALCLAWMSLAVL
CWDMALVRLLDRPRWRGWLQRRVGALDRLCGVLLLALGGWLVLAAL"

gene complement(8208. .9008)
/gene="PA5342"

CDS complement(8208. .9008)
/gene="PA5342"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG08727.1"
/db_xref="GI:9951661"
/translation="MAANRLLARPWLPGVELFHADFSGQAFGRHSHSAFAIGSILHGV
GGYQCRGRRHALPAGTLSLMNPEEPHTGHAESERLVYRMLYVEETRLLGLLGRKQLPR
GFGELNPADDGSVAAGLAWLAEDFQRSDALGLESRLLALLEQVFVRHGGLRPAAAAPR
DGGTTARLRDYLEAHATEALGLDQLAQLLQRHPRHLIEAFRRAYGVPPHTYQLQCRVR
QAKRLLLEERPLAELAQHLGFYDQAHFSTTFRRFTGVTPGQFRRSAKT"

gene complement(9008. .9859)
/gene="PA5343"

CDS complement(9008. .9859)

Query Match 100.0%; Score 24; DB 1; Length 13075;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccgttcgggcgcgaggaagcccgg 24
||||||||||||||||||||||||
Db 595 CCGTTCGGGCGCGAGGAAGCCCGG 572

COMP

RESULT 3
LMFLCHR16_03/c
WPCOMMENT
Sequence split into 11 fragments LOCUS LMFLCHR16 Accession AL499619

Fragment Name	Begin	End
LMFLCHR16_00	1	110000
LMFLCHR16_01	100001	210000
LMFLCHR16_02	200001	310000
LMFLCHR16_03	300001	410000
LMFLCHR16_04	400001	510000
LMFLCHR16_05	500001	610000

```
     LMFLCHR16_06        600001    710000
     LMFLCHR16_07        700001    810000
     LMFLCHR16_08        800001    910000
     LMFLCHR16_09        900001   1010000
     LMFLCHR16_10       1000001   1030105
Continuation (4 of 11) of LMFLCHR16 from base 300001 (AL499619 Leishmania major chromosc


  Query Match             86.7%;  Score 20.8;  DB 2;  Length 110000;
  Best Local Similarity   91.7%;  Pred. No. 97;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcccgg 24
          |||||| |||||||| ||||||||
Db  16542 CCGTTCCGGCGCGAGAAAGCCCGG 16519


RESULT   4
RME603644/c
LOCUS       RME603644             315000 bp    DNA     linear   BCT 16-AUG-2001
DEFINITION  Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymB;
            segment 3/6.
ACCESSION   AL603644 AL591985
VERSION     AL603644.1  GI:15140405
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 315000)
  AUTHORS   Finan,T.M., Weidner,S., Wong,K., Buhrmester,J., Chain,P.,
            Vorholter,F.J., Hernandez-Lucas,I., Becker,A., Cowie,A., Gouzy,J.,
            Golding,B. and Puhler,A.
  TITLE     From the Cover: The complete sequence of the 1,683-kb pSymB
            megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti
  JOURNAL   Proceedings of the National Academy of Sciences of the United
            States of America. 98 (17), 9889-9894 (2001)
   PUBMED   11481431
  REMARK    epub ahead of print
REFERENCE   2  (bases 1 to 315000)
  AUTHORS   Weidner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
            Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
            Germany
COMMENT     Submitted on behalf of Universitaet Bielefeld, Biologie IV
            (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and
            Department of Biology, McMaster University, 1280 Main Street West,
            Hamilton, Ontario, L8S 4K1 Canada
            mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE
            pEXO, pSymB.
FEATURES             Location/Qualifiers
     source          join(1. .104488,join(105027. .106910,join(107415. .229140,
                     230460. .315000)))
                     /organism="Sinorhizobium meliloti"
                     /plasmid="pSymB"
                     /strain="1021"
                     /db_xref="taxon:382"
                     /focus
     gene            complement(52. .405)
                     /gene="SMb20542"
     CDS             complement(52. .405)
                     /gene="SMb20542"
                     /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
                     functions"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative  protein in ISRm14"
                     /protein_id="CAC48921.1"
                     /db_xref="GI:15140406"
                     /translation="MIVAGQRLPILIATRPVDFRCGHQALALMVQTELKLDPHSGVTV
                     IFRSKRGDRLKILVWDGTGMVLTYKILEHGSFAWPKVQDGTMRLSRGQYEALFEGLDW
                     RRVMAQRVTAPSAAG"
     gene            289. .705
                     /gene="SMb20544"
     CDS             289. .705
                     /gene="SMb20544"
                     /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
                     functions"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative  protein in ISRm14"
                     /protein_id="CAC48922.1"
                     /db_xref="GI:15140407"
                     /translation="MRVELQLGLHHQSQRLMPTAEVHRPRCNQDRQSLARDDHTGARM
                     ARTRAATRSTGTSPGTRNTTSGPISNVKTDETASDSGGAKTFTASLGSAAIAKGTNAG
                     SASPLSDRPGIASNGSARMPSRACRRQKESWCGTTS"
     gene            complement(402. .845)
                     /gene="SMb20543"
     CDS             complement(402. .845)
                     /gene="SMb20543"
                     /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
                     functions"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative  protein in ISRm14"
                     /protein_id="CAC48923.1"
                     /db_xref="GI:15140408"
                     /translation="MAGDGFVGRYEVVEPRRGNRRWPDAVKARIVAESLEPGVRVVDV
                     ARRHDVVPHQLSFWRRQAREGILALPFEAMPGLSESGDAEPAFVPLAIAAEPSEAVNV
                     LAPPLSEAVSSVLTLEIGPDVVLRVPGDVPVERVAALVRAMRAPV"
     gene            complement(1438. .1656)
                     /gene="SMb20545"
     CDS             complement(1438. .1656)
                     /gene="SMb20545"
                     /function="MISCELLANEOUS; Hypothetical/Partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC48924.1"
                     /db_xref="GI:15140409"
                     /translation="MDDREAGSAADAEPAPDLSEVAAQCAAADLLDFAAQLAGLAEDL
                     KALAARPIEPTPTLAEPDHPPEATSSDL"
     gene            complement(1843. .2361)
                     /gene="SMb20546"
     CDS             complement(1843. .2361)
                     /gene="SMb20546"
                     /function="MISCELLANEOUS; Hypothetical/Global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="CONSERVED HYPOTHETICAL PROTEIN"
                     /protein_id="CAC48925.1"
                     /db_xref="GI:15140410"
                     /translation="MAEKWDECVIIDLPNLEGIQIVWSPAYAAQLLSENWPDTTGRSY
                     SAALNSCTDAMFGAASAAPARDAFLAAVEEARVKTLRWHHAAPPLVKGNDSIEGGYVA
                     ELGFQARSHSVRAWSTGGGRQADEFSARSSLRSSESGAVLVSVPSTDARLISFTLPFA
                     ASRTCRAVSAHW"
     gene            complement(2518. .2796)
                     /gene="SMb20547"
     CDS             complement(2518. .2796)
                     /gene="SMb20547"
```

```
                /function="MISCELLANEOUS; Hypothetical/Partial homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="HYPOTHETICAL PROTEIN"
                /protein_id="CAC48926.1"
                /db_xref="GI:15140411"
                /translation="MYEPFEAQIKAWLKEGPTLSAAGVLRRLTEIDPIRFTEKNIRMV
                QMLVKAWRMEMAVLVILDGGWIKSLPSAAAAAVDTGTSASAVLGNILR"
     gene       3930. .4424
                /gene="SMb20548"
     CDS        3930. .4424
                /gene="SMb20548"
                /function="MISCELLANEOUS; Unknown"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="HYPOTHETICAL PROTEIN"
                /protein_id="CAC48927.1"
                /db_xref="GI:15140412"
                /translation="MESIIERNCRKDGFDIFMPSFYTELRHHRTKQILQKRFPFLVGY
                GFVSLPRLNFEELRRVDGVVCFLRGANYGPLEFPDGYHRGPVLRRSRATAGLPLQAAL
                PEGDRAPRANPALGRPAPEDPAKGSESARLHGRPCRECYRFSKPADQRASADNYQWIE
                QPHC"
     gene       6048. .6365
                /gene="SMb20549"
     CDS        6048. .6365
                /gene="SMb20549"
                /function="MISCELLANEOUS; Hypothetical/Global homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="CONSERVED HYPOTHETICAL PROTEIN"
                /protein_id="CAC48928.1"
                /db_xref="GI:15140413"
                /translation="MRALRKASAHYDVMSAEQIRALPVGQLASMNCLIYSRATAPHLA
                FTVECLKAWGFEYKSFMAWRKTTAAGKVRMGTGYRVRTTGEIVFVGTLGNPKQSHVPP
                TIS"
     gene       complement(6533. .6778)
                /gene="SMb20550"
     CDS        complement(6533. .6778)
                /gene="SMb20550"
                /function="MISCELLANEOUS; Hypothetical/Global homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="CONSERVED HYPOTHETICAL PROTEIN"
                /protein_id="CAC48929.1"
                /db_xref="GI:15140414"
                /translation="MGTVWGECVILDLPDLDGVQIVWHPAYAARLLIDSWPVDHGRAY
                GTAINACADAMVGICTGERARIAFIAAINEAKIDLLA"
     gene       complement(6952. .7398)
                /gene="SMb20551"
     CDS        complement(6952. .7398)
                /gene="SMb20551"
                /function="MISCELLANEOUS; Unknown"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="HYPOTHETICAL PROTEIN"
                /protein_id="CAC48930.1"
                /db_xref="GI:15140415"
                /translation="MVMAHLGRVATKIDHRPLELAEKRQQINTGRWSAAPSLLNTVVF
```

```
  Query Match             80.8%;  Score 19.4;  DB 1;  Length 315000;
  Best Local Similarity   95.2%;  Pred. No. 2.7e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 gttcgggcgcgaggaagcccg 23
           ||||| |||||||||||||||
Db 240027 GTTCGAGCGCGAGGAAGCCCG 240007


RESULT   5
SC9A10/c
LOCUS       SC9A10                  9000 bp    DNA     linear   BCT 11-AUG-1998
DEFINITION  Streptomyces coelicolor cosmid 9A10.
ACCESSION   AL031260
VERSION     AL031260.1  GI:3413819
KEYWORDS    16S rRNA gene; dapA; dapB; dihydrodipicolinate reductase;
            dihydrodipicolinate synthase; gpsI, guanosine pentaphosphate
            synthetase; polyribonucleotide nucleotidyltransferase; protease.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 9000)
  AUTHORS   Harris,D. and Taylor,K.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9000)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE   3  (bases 1 to 9000)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 9A10 lies
            between 3C3 and 7C7 on the AseI-B genomic restriction fragment.
```

```
FEATURES             Location/Qualifiers
     source          1. .9000
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 9A10"
     gene            1. .175
                     /gene="gpsI"
     CDS             <1. .175
                     /gene="gpsI"
                     /note="SC9A10.01, gpsI, guanosine pentaphosphate
                     synthetase/ polyribonucleotide nucleotidyltransferase,
                     partial CDS, len: >57 aa; almost identical to Streptomyces
                     antibioticus gpsI TR:Q53597 (EMBL:U19858) guanosine
                     pentaphosphate synthetase (740 aa), 87.9% identity in 58
                     aa overlap. Overlaps and extends SC3C3.23."
                     /codon_start=2
                     /transl_table=11
                     /product="guanosine pentaphosphate synthetase/
                     polyribonucleotide nucleotidyltransferase"
                     /protein_id="CAA20288.1"
                     /db_xref="GI:3413820"
                     /db_xref="SPTREMBL:O86834"
                     /translation="IRKLAGGKRVENVEDVLGVGAKVQVEIAEIDSRGKLSLIPVIEG
                     EAGDDDKKDDADK"
     misc_feature    1. .104
                     /gene="gpsI"
                     /note="nominal overlap with 3C3 from: 31279 to: 31382"
     RBS             155. .159
                     /gene="gpsI"
                     /note="possible RBS upstream of SC9A10.02"
     gene            172. .1551
                     /gene="SC9A10.02"
     CDS             172. .1551
                     /gene="SC9A10.02"
                     /note="SC9A10.02, probable protease, len: 459 aa; similar
                     to members of the insulinase family e.g. YMXG_BACSU
                     hypothetical processing protease (409 aa), fasta scores;
                     opt: 648 z-score: 1449.2 E(): 0, 39.7% identity in 413 aa
                     overlap and MPP2_YEAST mitochondrial processing peptidase
                     beta subunit (462 aa), fasta scores; opt: 583 z-score:
                     864.6 E(): 0, 29.7% identity in 448 aa overlap.
                     Alternative start possible at aa 36. Contains PS00143
                     Insulinase family, zinc-binding region signature and Pfam
                     match to entry PF00675 Insulinase, Insulinase (proteinase
                     M16), score 171.30, E-value 1.6e-47"
                     /codon_start=1
                     /transl_table=11
                     /product="putative protease"
                     /protein_id="CAA20289.1"
                     /db_xref="GI:3413821"
                     /db_xref="SPTREMBL:O86835"
                     /translation="MTSRSATATARTSSEARAVARTQTLIKGEHGIGTVRRTTLPGGL
                     RIVTETLPSVRSATFGIWAHVGSRDETPALNGATHYLEHLLFKGTRKRSALDISSAID
                     AVGGEMNAFTAKEYTCYYARVLDTDLPLAIDVVCDMLTGSLIQEEDVDVERGAILEEI
                     AMTEDDPGDCVHDLFAHTMFGDNALGRPVLGTVDTVNALTADRIRRFYRKHYDPTHLV
                     VAAAGNVDHNKVVRQVRAAFEKSGALKDPAAQPLAPRAGRRTVRAAGRVELIGRKTEQ
                     AHVILGMPGLARTDERRWAMGVLNTALGGGMSSRLFQEVREKRGLAYSVYSYTSGFAD
                     CGLFGVYAGCRPSQVHDVLKICRDELDHVAEHGLTDDEIGRAVGQLQGSTVLGLEDTG
                     ALMNRIGKSELCWGEQMSVDDMLARIASVTPDDVRAVARDVLGRRPSLSVIGPLKDKQ
                     ASRLHDAVA"
     stem_loop       192. .231
                     /gene="SC9A10.02"
                     /note="hairpin loop with 18 /19 bp stem"
     misc_feature    301. .750
                     /gene="SC9A10.02"
                     /note="Pfam match to entry PF00675 Insulinase, Insulinase
                     (proteinase M16), score 171.30, E-value 1.6e-47"
     misc_feature    367. .438
                     /gene="SC9A10.02"
                     /note="PS00143 Insulinase family, zinc-binding region
                     signature"
     RBS             1562. .1566
                     /note="possible RBS upstream of dapB"
     gene            1577. .2329
                     /gene="dapB"
     CDS             1577. .2329
                     /gene="dapB"
                     /note="SC9A10.03, dapB, dihydrodipicolinate reductase,
                     len: 250 aa; highly similar to many e.g. DAPB_CORGL
                     dihydrodipicolinate reductase (EC 1.3.1.26) (248 aa),
                     fasta scores; opt: 946 z-score: 1333.7 E(): 0, 56.1%
                     identity in 246 aa overlap. Contains PS01298
                     Dihydrodipicolinate reductase signature"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dihydrodipicolinate reductase"
                     /protein_id="CAA20290.1"
                     /db_xref="GI:3413822"
                     /db_xref="SPTREMBL:O86836"
                     /translation="MSKLRVAVLGAKGRIGSEAVRAVEAAEDMELVAALGRGDGLEAL
                     AESGAQVAVELTTPASVMDNLDYCLRHGIHAVVGTTGWTDERLARLNAWLDASPGTGV
                     LIAPNFSIGAVLTMKFAQIAAPYFESVEVVELHHPNKVDAPSGTATRTAQLIAQARQK
                     AGSAPAPDATATALDGARGANVDGVPVHAVRLRGLLAHQEVLLGAEGETLTVRHDSLH
                     HSSFMPGILLGARRVVTTPGLTFGLEHFLDLN"
     misc_feature    1964. .2017
                     /gene="dapB"
                     /note="PS01298 Dihydrodipicolinate reductase signature"
     gene            2343. .2798
                     /gene="SC9A10.04"
     CDS             2343. .2798
                     /gene="SC9A10.04"
                     /note="SC9A10.04, possible membrane protein, len: 151 aa;
                     similar to hypothetical proteins from Mycobacterium bovis
                     and tuberculosis that follow dapB, e.g. YDAB_MYCBO
                     hypothetical 19.3 kd protein in dapb 3'region (177 aa),
                     fasta scores; opt: 377 z-score: 696.1 E(): 1.7e-31, 47.1%
                     identity in 136 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="CAA20291.1"
                     /db_xref="GI:3413823"
                     /db_xref="SPTREMBL:O86837"
                     /translation="MRAKITYLVTAAVLVFYFLLVGSRGVLLIKTGTLLTVTFGVAVL
                     ILPVIGLWFLWKNTQFVRRANQLAAELDAEGGLPVDELKRTPSGRVDRESADAVFALR
                     RAETEDAPGDWRSWFRLAVAYHDARDTPRARKAMQRAIALHDGKHVEAA"
     stem_loop       2807. .2864
                     /note="hairpin loop with 29 bp stem"
     misc_feature    2827. .2832
                     /note="single clone coverage only"
     gene            complement(2874. .3428)
                     /gene="SC9A10.05c"
     CDS             complement(2874. .3428)
                     /gene="SC9A10.05c"
                     /note="SC9A10.05c, probable membrane protein, len: 184 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="CAA20292.1"
                     /db_xref="GI:3413824"
                     /db_xref="SPTREMBL:O86838"
                     /translation="MPLPFLTADRAFDTADDVALPYDDRDLWRRPYRPGPWRVGMAAL
                     MLLLASFVLFAAVIIAATGGLSSAGVVFGIAVVVILCALRMLRMGAWVSAHGVRQVGF
                     FTTRTVAWQQVASLRTVQQPVRWLGLPRTVQGQALLLVRQGRSAEPLSPLLTSHNSDF
                     LARDGAFNRAADAVEAWADEYRRG"
     RBS             complement(3435. .3438)
                     /note="possible RBS upstream of SC9A10.05c"
     gene            complement(3545. .3784)

  Query Match           80.0%;  Score 19.2;  DB 1;  Length 9000;
  Best Local Similarity  87.5%;  Pred. No. 6.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ccgttcgggcgcgaggaagcccgg 24
           ||| ||||||| || |||||||||
```

```
Db    6644 CCGCTCGGGCGGGACGAAGCCCGG 6621


RESULT   6
AC005336
LOCUS       AC005336               36874 bp    DNA     linear   PRI 15-JUN-2001
DEFINITION  Homo sapien chromosome 19, cosmid F20129, complete sequence.
ACCESSION   AC005336
VERSION     AC005336.1  GI:3347821
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 36874)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
            Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 1.5 Mb OLFR-rich region in 19p13.1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36874)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Cosmid
            F20129 overlaps cosmid R26420 (AC004791) to the left from bases 1
            to 703, and is expected to overlap cosmid R28342 to the right by
            approx. 6 kb. Additional map and sequence information are available
            at: http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1. .36874
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19p13.1 from OLFR to MEL"
                     /clone="F20129"
                     /cell_line="UV5HL9-5B"
                     /clone_lib="LL19NC02 F chromosome 19 specific cosmid"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from human-hamster hybrid UV5HL9-5B, which
                     carries chromosome 19 as its only human chromosome."
     misc_feature    190. .362
                     /note="DDS similarity to AA377259 EST89796 Small intestine
                     II Homo sapiens cDNA 3' end similar to cytochrome P450,
                     subfamily IVF, polypeptide 2; Score: 334  Identity:
                     170/173 (98%)."
     mRNA            complement(260. .20308)
                     /product="Human cytochrome P450 4F2 (CYP4F2) mRNA"
     repeat_region   complement(363. .963)
                     /rpt_family="L1MB5"
     gene            complement(1007. .19847)
                     /gene="CYP4F2"
                     /note="cytochrome P450 4F2"
     CDS             complement(join(1007. .1172,1582. .1664,1840. .1904,
                     2000. .2133,8160. .8289,8478. .8544,11659. .11929,
                     12548. .12669,14545. .14672,14765. .14818,17742. .17886,
                     19650. .19847))
                     /gene="CYP4F2"
                     /note="cytochrome P450 4F2; leukotriene B4
                     omega-hydroxylase; LEUKOTRIENE-B4 20-MONOOXYGENASE;
                     YTOCHROME P450-LTB-OMEGA; LEUKOTRIENE-B4
                     OMEGA-HYDROXYLASE"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="CYP4F2"
                     /protein_id="AAC27730.1"
                     /db_xref="GI:3347822"
                     /translation="MSQLSLSWLGLWPVAASPWLLLLLVGASWLLAHVLAWTYAFYDN
                     CRRLRCFPQPPRRNWFWGHQGMVNPTEEGMRVLTQLVATYPQGFKVWMGPISPLLSLC
                     HPDIIRSVINASAAIAPKDKFFYSFLEPWLGDGLLLSAGDKWSRHRRMLTPAFHFNIL
                     KPYMKIFNESVNIMHAKWQLLASEGSACLDMFEHISLMTLDSLQKCVFSFDSHCQEKP
                     SEYIAAILELSALVSKRHHEILLHIDFLYYLTPDGQRFRRACRLVHDFTDAVIQERRR
                     TLPSQGVDDFLQAKAKSKTLDFIDVLLLSKDEDGKKLSDEDIRAEADTFMFEGHDTTA
                     SGLSWVLYHLAKHPEYQERCRQEVQELLKDREPKEIEWDDLAHLPFLTMCMKESLRLH
                     PPVPVISRHVTQDIVLPDGRVIPKGIICLISVFGTHHNPAVWPDPEVYDPFRFDPENI
                     KERSPLAFIPFSAGPRNCIGQTFAMAEMKVVLALTLLRFRVLPDHTEPRRKPELVLRA
                     EGGLWLRVEPLS"
     repeat_region   2583. .2871
                     /rpt_family="AluY"
     repeat_region   3149. .3382
                     /rpt_family="L1MD2"
     repeat_region   3394. .3472
                     /rpt_family="L1MA7"
     repeat_region   3600. .3711
                     /rpt_family="(TAGA)n"
     repeat_region   3712. .3841
                     /rpt_family="(TAGA)n"
     repeat_region   4000. .4198
                     /rpt_family="L1MA7"
     repeat_region   4199. .4500
                     /rpt_family="AluSg"
     repeat_region   4515. .4818
                     /rpt_family="L1MA7"
     repeat_region   4818. .4962
                     /rpt_family="L1MA7"
     repeat_region   5283. .6013
                     /rpt_family="L1MA4"
     repeat_region   6015. .6314
                     /rpt_family="AluSg"
     repeat_region   6317. .6350
                     /rpt_family="(GA)n"
     repeat_region   6359. .6866
                     /rpt_family="L1MA4"
     repeat_region   7064. .7329
                     /rpt_family="LINE2"
     repeat_region   complement(7330. .7392)
                     /rpt_family="(GA)n"
     repeat_region   complement(9042. .9092)
                     /rpt_family="(GA)n"
     repeat_region   9378. .9500
                     /rpt_family="(CA)n"
     repeat_region   9554. .9678
                     /rpt_family="(CA)n"
     repeat_region   9985. .10104
                     /rpt_family="(CA)n"
     repeat_region   10176. .10292
                     /rpt_family="(CA)n"
     repeat_region   10393. .10512
                     /rpt_family="(CAAA)n"
     repeat_region   10550. .10668
                     /rpt_family="(CA)n"
     repeat_region   10670. .10744
                     /rpt_family="(CA)n"
     repeat_region   10746. .10916
                     /rpt_family="(CA)n"
     repeat_region   10919. .11035
                     /rpt_family="(CA)n"
     repeat_region   12968. .13227
                     /rpt_family="MIR"
     repeat_region   complement(14390. .14462)
                     /rpt_family="MIR"
     repeat_region   complement(16352. .16423)
                     /rpt_family="MIR"
     repeat_region   17916. .17998
                     /rpt_family="MLT2D"
     repeat_region   complement(18011. .18148)
                     /rpt_family="(TAGA)n"
     repeat_region   complement(18193. .18379)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(18372. .18512)
```

```
                     /rpt_family="(TAGA)n"
     repeat_region   20478. .20676
                     /rpt_family="(TAGA)n"
     repeat_region   20678. .20737
                     /rpt_family="(TAGA)n"
     repeat_region   20909. .20950
                     /rpt_family="POLY_A"
     repeat_region   complement(21030. .21066)
                     /rpt_family="(TGG)n"
     misc_feature    complement(21037. .21130)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 82.000"
     misc_feature    complement(21374. .21496)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 80.000"
     repeat_region   21382. .21506
                     /rpt_family="pTR5"
     repeat_region   complement(21508. .21550)
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(21555. .21761)
                     /rpt_family="L1PA2"
     repeat_region   21761. .21852
                     /rpt_family="L1PA2"
     misc_feature    complement(21844. .22336)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 57.000"
     repeat_region   21871. .22548
                     /rpt_family="pTR5"
     repeat_region   22549. .22624
                     /rpt_family="MER51A"
     repeat_region   22625. .22856
                     /rpt_family="pTR5"
     misc_feature    22855. .23067
                     /note="BLASTN similarity to AQ036765 (315. .527); match:
                     0.98, score: 6.1e-79; database searched: month.na;
                     CIT-HSP-2331C14.TF CIT-HSP Homo sapiens genomic clone
                     2331C14"
     repeat_region   23538. .23909
                     /rpt_family="MSTA"
     repeat_region   complement(23928. .24014)
                     /rpt_family="AluJ/FLAM"
     repeat_region   24015. .24053
                     /rpt_family="MSTA"
     misc_feature    complement(24562. .25475)
                     /note="Mitochondrial RNA polymerase pseudogene, partial
                     CDS:
                     CNGLQHYAALGRDSVGAASVNLEPSDVPQDVYSGVAAQVEVFRRQDAQRGMRVAQVLE
                     AFITRKVVKQTVMTVVYGVTRYGGRLQIEKRLRELNDFPQEFVWEASHYLVRQVFRSL
                     QEMFSGTRAIQHWLIESAHLVSHTGSVVEWVTPLGVPIIQPYRLESKVKQSGGGIQSI
                     TYTHNRDISRKPNTPKQKNGFRPNFIHSLDSSHMMLTTLHCYRKGLTFVSVHDCYRTH
                     AADVSIMNQVCREQFACTASPSCRTCPGSWSGGSALSPRRSWRPAR"
     repeat_region   25476. .25744
                     /rpt_family="L1"
     repeat_region   complement(25752. .26229)
                     /rpt_family="L1M2"
     repeat_region   26230. .26680
                     /rpt_family="MLT2A"
     repeat_region   26706. .26755
                     /rpt_family="MLT2A"

  Query Match             78.3%;  Score 18.8;  DB 9;  Length 36874;
  Best Local Similarity   90.9%;  Pred. No. 7.2e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 gttcgggcgcgaggaagcccgg 24
           | ||||||||||||| ||||||
Db   34151 GCTCGGGCGCGAGGAGGCCCGG 34172


RESULT   7
AC020950/c
LOCUS       AC020950               39877 bp    DNA     linear   PRI 17-OCT-2001
DEFINITION  Homo sapiens chromosome 19 clone LLNLR-244B2, complete sequence.
ACCESSION   AC020950
VERSION     AC020950.6  GI:16195213
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 39877)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39877)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 39877)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 39877)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 17, 2001 this sequence version replaced gi:8844113.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 98.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
            STS Content:
            SHGC-105463 G58686.
FEATURES             Location/Qualifiers
     source          1. .39877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="LLNLR-244B2"
BASE COUNT    10653 a    8957 c    8867 g   11400 t
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 9;  Length 39877;
  Best Local Similarity   90.9%;  Pred. No. 7.1e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 gttcgggcgcgaggaagcccgg 24
           | ||||||||||||| ||||||
Db   23133 GCTCGGGCGCGAGGAGGCCCGG 23112


RESULT   8
AL162853
LOCUS       AL162853               123693 bp    DNA     linear   PRI 13-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-363G2 on chromosome 13, complete
            sequence.
ACCESSION   AL162853
VERSION     AL162853.17  GI:14456190
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123693)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 14, 2001 this sequence version replaced gi:14141275.
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr13
RP11-363G2 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-363G2 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-124N19 is at 123594 in this sequence. The true right end of clone RP11-274P12 is at 100 in this sequence.

FEATURES Location/Qualifiers

source	1. .123693
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="13"
	/clone="RP11-363G2"
	/clone_lib="RPCI-11.2"
repeat_region	146. .455
	/note="AluSx repeat: matches 1. .310 of consensus"
repeat_region	1159. .1208
	/note="25 copies 2 mer tt 74% conserved"
repeat_region	1738. .2031
	/note="AluSp repeat: matches 11. .306 of consensus"
repeat_region	2845. .2890
	/note="23 copies 2 mer ca 73% conserved"
repeat_region	3907. .4053
	/note="MIR repeat: matches 21. .176 of consensus"
repeat_region	4767. .5022
	/note="L1M4 repeat: matches 2752. .3022 of consensus"
repeat_region	5459. .5752
	/note="AluY repeat: matches 1. .293 of consensus"
repeat_region	5763. .5869
	/note="L1ME2 repeat: matches 6004. .6110 of consensus"
repeat_region	5876. .5938
	/note="MLT1A1 repeat: matches 161. .226 of consensus"
repeat_region	6150. .6910
	/note="MER52A repeat: matches 274. .1752 of consensus"
repeat_region	7127. .7404
	/note="AluJo repeat: matches 1. .281 of consensus"
repeat_region	7652. .7748
	/note="L2 repeat: matches 2600. .2702 of consensus"
repeat_region	7866. .8037
	/note="MIR repeat: matches 21. .204 of consensus"
repeat_region	10436. .11460
	/note="L2 repeat: matches 1410. .2698 of consensus"
repeat_region	12866. .13180
	/note="AluSq repeat: matches 1. .312 of consensus"
repeat_region	13593. .13628
	/note="18 copies 2 mer gt 86% conserved"
repeat_region	13858. .14022
	/note="MLT1C repeat: matches 77. .248 of consensus"
repeat_region	14055. .15039
	/note="L1MC2 repeat: matches 4378. .5361 of consensus"
repeat_region	15040. .15406
	/note="THE1C repeat: matches 1. .369 of consensus"
repeat_region	15407. .16272
	/note="L1MC2 repeat: matches 5361. .6241 of consensus"
repeat_region	16275. .16300
	/note="13 copies 2 mer tg 96% conserved"
repeat_region	16325. .16413
	/note="L1MC2 repeat: matches 6239. .6325 of consensus"
repeat_region	16397. .16799
	/note="L1MB8 repeat: matches 5297. .5709 of consensus"
repeat_region	16800. .17112
	/note="AluY repeat: matches 1. .308 of consensus"
repeat_region	17113. .17563
	/note="L1MB8 repeat: matches 5709. .6129 of consensus"
repeat_region	17666. .17956
	/note="AluY repeat: matches 1. .297 of consensus"
repeat_region	18501. .18815
	/note="AluY repeat: matches 1. .307 of consensus"
repeat_region	18860. .19842
	/note="L1MA8 repeat: matches 4857. .5769 of consensus"
repeat_region	19843. .20151
	/note="AluSg repeat: matches 1. .309 of consensus"
repeat_region	20152. .20466
	/note="L1MA8 repeat: matches 5769. .6088 of consensus"
repeat_region	20975. .21271
	/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region	21591. .21894
	/note="AluJb repeat: matches 5. .297 of consensus"
repeat_region	21918. .22218
	/note="AluY repeat: matches 7. .307 of consensus"
misc_feature	complement(22425. .22714)
	/note="match: GSS: Em:B89882"
repeat_region	22560. .22654
	/note="AluY repeat: matches 217. .311 of consensus"
repeat_region	26136. .26564
	/note="L1MB8 repeat: matches 5752. .6173 of consensus"
repeat_region	26569. .26879
	/note="L1MB8 repeat: matches 5404. .5722 of consensus"
repeat_region	27823. .27886
	/note="L2 repeat: matches 2679. .2741 of consensus"
repeat_region	28146. .28511
	/note="L2 repeat: matches 1572. .1949 of consensus"
repeat_region	28603. .29019
	/note="MLT1C repeat: matches 3. .460 of consensus"
repeat_region	30976. .31289
	/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region	31350. .31405
	/note="28 copies 2 mer ca 69% conserved"
repeat_region	31569. .31854
	/note="AluSg repeat: matches 1. .285 of consensus"
repeat_region	32300. .32416
	/note="L1MB8 repeat: matches 5625. .5750 of consensus"
repeat_region	33065. .33222
	/note="L2 repeat: matches 2540. .2710 of consensus"
repeat_region	33577. .33858
	/note="L1MEc repeat: matches 1328. .1621 of consensus"
repeat_region	33947. .33959
	/note="L1MEc repeat: matches 1760. .1771 of consensus"
repeat_region	33960. .34166
	/note="HAL1 repeat: matches 697. .914 of consensus"
repeat_region	34167. .34178
	/note="L1MEc repeat: matches 1771. .2002 of consensus"
repeat_region	34290. .34533
	/note="L1MC/D repeat: matches 5321. .5558 of consensus"
repeat_region	34633. .34700
	/note="MER44 repeat: matches 506. .574 of consensus"
repeat_region	34758. .34851
	/note="L1MA7 repeat: matches 6136. .6229 of consensus"
repeat_region	34852. .35438
	/note="L1PA16 repeat: matches 5585. .6157 of consensus"
repeat_region	35439. .36300

```
                /note="L1MA7 repeat: matches 5273. .6137 of consensus"
     misc_feature    37701. .38100
                /note="match: GSS: Em:AQ152136"
     misc_feature    38805. .38904
                /note="match: GSS: Em:AQ197787"
     misc_feature    complement(38816. .38907)
                /note="match: GSS: Em:AQ793043"
     repeat_region   39006. .39285
                /note="AluSc repeat: matches 1. .280 of consensus"
     repeat_region   39322. .39636
                /note="AluSq repeat: matches 1. .313 of consensus"
     repeat_region   39657. .39714
                /note="MIR repeat: matches 57. .115 of consensus"
     repeat_region   40171. .40321
                /note="LTR33 repeat: matches 1. .151 of consensus"
     repeat_region   41075. .41201
                /note="AluJb repeat: matches 170. .296 of consensus"
     repeat_region   41202. .41429
                /note="LTR1 repeat: matches 225. .492 of consensus"
     repeat_region   41430. .41653
                /note="AluY repeat: matches 1. .220 of consensus"
     repeat_region   41656. .41777
                /note="L1M4 repeat: matches 4611. .4735 of consensus"
     repeat_region   41738. .42963
                /note="L1MA8 repeat: matches 5042. .6290 of consensus"
     repeat_region   42996. .44070
                /note="L1ME3 repeat: matches 4694. .5818 of consensus"
     repeat_region   44071. .44248
                /note="AluSq repeat: matches 142. .311 of consensus"
     repeat_region   44249. .44577
                /note="AluSp repeat: matches 1. .312 of consensus"
     repeat_region   44578. .44716
                /note="AluSq repeat: matches 1. .142 of consensus"
     repeat_region   44717. .44845
                /note="L1ME3 repeat: matches 5818. .6059 of consensus"
     repeat_region   44957. .45809
                /note="L1PA13 repeat: matches 5293. .6154 of consensus"

  Query Match             78.3%;  Score 18.8;  DB 9;  Length 123693;
  Best Local Similarity   90.9%;  Pred. No. 5.7e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gttcgggcgcgaggaagcccgg 24
          | ||||||||||||| ||||||
Db  73573 GCTCGGGCGCGAGGAGGCCCGG 73594


RESULT  9
AC016396
LOCUS       AC016396              151764 bp    DNA     linear   PRI 15-JUN-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-179B15, complete sequence.
ACCESSION   AC016396
VERSION     AC016396.5  GI:7715613
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151764)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151764)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 151764)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 151764)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On May 7, 2000 this sequence version replaced gi:6910512.
FEATURES             Location/Qualifiers
     source          1. .151764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-179B15"
                     /clone_lib="RPCI-11"
BASE COUNT    47476 a  29555 c  29458 g  45274 t      1 others
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 9;  Length 151764;
  Best Local Similarity   90.9%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gttcgggcgcgaggaagcccgg 24
          | ||||||||||||| ||||||
Db  69869 GCTCGGGCGCGAGGAGGCCCGG 69890


RESULT  10
AL358155
LOCUS       AL358155              182433 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 10 clone RP11-637J22, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL358155
VERSION     AL358155.17  GI:14529817
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182433)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 22, 2001 this sequence version replaced gi:14348459.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA637J22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 182127 bases at least Q40
            Consensus quality: 182344 bases at least Q30
            Consensus quality: 182368 bases at least Q20
            Insert size: 182433; sum-of-contigs
            Insert size: 201690; 10.2% error; agarose-fp
            Quality coverage: 7.30x in Q20 bases; sum-of-contigs Quality
            coverage: 6.60x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .182433
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
```

```
                     /clone="RP11-637J22"
                     /clone_lib="RPCI-11.3"
     misc_feature    1. .182433
                     /note="assembly_fragment:00033
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    58805 a  35233 c  34745 g  53650 t
ORIGIN


  Query Match             78.3%;  Score 18.8;  DB 2;  Length 182433;
  Best Local Similarity   90.9%;  Pred. No. 5.2e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gttcgggcgcgaggaagcccgg 24
          | |||||||||||||| ||||||
Db 133774 GCTCGGGCGCGAGGAGGCCCGG 133795


RESULT  11
AC077474/c
ID   AC077474   standard; DNA; HTG; 960 BP.
XX
AC   AC077474;
XX
SV   AC077474.1
XX
DT   03-AUG-2000 (Rel. 64, Created)
DT   03-AUG-2000 (Rel. 64, Last updated, Version 1)
XX
DE   Giardia intestinalis clone MJ4582 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-960
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-960
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (02-AUG-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *        1       960: contig of 960 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .960
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="MJ4582"
XX
SQ   Sequence 960 BP; 220 A; 248 C; 226 G; 263 T; 3 other;


  Query Match             75.8%;  Score 18.2;  DB 33;  Length 960;
  Best Local Similarity   87.0%;  Pred. No. 2.6e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcccg 23
          ||||  |||||||||||||||| |
Db    349 CCGTCAGGGCGCGAGGAAGCCTG 327


RESULT  12
AF043641/c
LOCUS       AF043641               2214 bp    mRNA    linear   VRT 18-JUL-2000
DEFINITION  Barbus barbus p73 mRNA, complete cds.
ACCESSION   AF043641
VERSION     AF043641.1  GI:4689085
KEYWORDS    .
SOURCE      barbel.
  ORGANISM  Barbus barbus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Barbus.
REFERENCE   1  (bases 1 to 2214)
  AUTHORS   Bhaskaran,A., May,D., Rand-Weaver,M. and Tyler,C.R.
  TITLE     Molecular characterization of the first non-mammalian p73 cDNA
  JOURNAL   Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 126 (1), 49-57
            (2000)
  MEDLINE   20362516
REFERENCE   2  (bases 1 to 2214)
  AUTHORS   Bhaskaran,A., May,D., Rand-Weaver,M. and Tyler,C.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1998) Biology and Biochemistry, Brunel
            University, Kingston Lane, Uxbridge, Middx UB8 3PH, UK
FEATURES             Location/Qualifiers
     source          1. .2214
                     /organism="Barbus barbus"
                     /db_xref="taxon:40830"
     CDS             125. .2050
                     /codon_start=1
                     /product="p73"
                     /protein_id="AAD27752.1"
                     /db_xref="GI:4689086"
                     /translation="MSQSSTADEGTTFEHLWSTLEPDSTYFELPQAGHSGDRVASSSL
                     PSNRAEVCMDVYHMRDMRDMNDNVMSQYSLLSSSMEQGLGNRAASTSPYSSETTSNVP
                     TPSPYSQPNSTFEAMSPAPAIPSNTDYPGPHNFEVTFQQSSTAKSATWTYSPLLKKLY
                     CQIAKTCPIQIKLASSPPNGSVIRAMPIYKKAEHVTEVVKRCPNHELGRDFNESQTAP
                     ASHLIRVEGNNLSQYVDDPVTGRQSALVPYEAPQVGTEFTTILYNFMCNSSCVGGMNR
                     RPILIIITLETRDGQVLGRRSFEGRICACPGRDRKADEDHFREQQALNESVAKNGNAN
                     KRNFKQTPGNITGPSINIKKGRHGEEEMYYIPVRGRENFDILMKIKDSLELVEYVPQH
                     FVDSYRQQQQQLLQRQSHVASPSPYGTLNNMNKIHGSISKLPSVNQLVTQQTQQSAGP
                     SASMSHMGTNMLGGHHMQSNGDVNGAHASQSLVSTSHCTPPPPYNPDPSLVSFLTSLG
                     CQNCIDYFTSQGLQSVYHLQTLSMEDLGALKIPEQFRMAIWRGLQDMKQGHDYGQQLI
                     RSSSNMATMAIGGGGELQRQRVMEAVHFRVRHTITIPNRGAPTGAEEWADFGFDMPDC
                     RIHKHSIKEEFAESDVH"
BASE COUNT      597 a    622 c    552 g    443 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 5;  Length 2214;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcccg 23
          ||||| ||| |||||||||||| |
Db    663 CCGTTTGGGGGCGAGGAAGCCAG 641


RESULT  13
LMFL673
LOCUS       LMFL673                21057 bp    DNA     linear   INV 23-DEC-1999
```

DEFINITION Leishmania major Friedlin chromosome 23 cosmid L673.
ACCESSION AL135898
VERSION AL135898.1 GI:6635092
KEYWORDS multidrug resistance protein, copy 1; multidrug resistance protein, copy 2.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE 1 (bases 1 to 21057)
AUTHORS Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
TITLE A physical map of the Leishmania major Friedlin genome
JOURNAL Genome Res. 8 (2), 135-145 (1998)
MEDLINE 98146435
REFERENCE 2 (bases 1 to 21057)
AUTHORS Zimmermann,W., Wambutt,R., Ivens,A.C., Murphy,L., Quail,M., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1999) European Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk) and Agowa GmbH, Glienickerweg 185, D-12489, Berlin, Germany
COMMENT see http://www.ebi.ac.uk/parasites/leish.html
Notes:
Details of leishmania sequencing at the Sanger Centre are available on the World Wide Web.
see http://www.sanger.ac.uk/Projects/L_major/
CDS are numbered using the following system eg L673.01. L673 (cosmid name), .01 (first CDS)
To make the cosmid library Leishmania major Friedlin DNA was partially digested with Sau3AI prior to cloning into BamHI site of the cosmid shuttle vector cLHYG (Ryan et al. 1993 Gene 131:145-150). The sequence of the packaged vector was determined by Peter Myler and Ken Stuart at Seattle Biomedical Research Institute, and is available as accession number U59231.
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

Gene prediction is done using:
(1)
the FramePlot program of Bibb et al.,
Gene 30:157-166(1984) as implemented
at http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. (2)
codon preference based on the codon usage table for Leishmania at
http://www.kazusa.or.jp/codon/
(3)
the Hexamer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C.elegans Genome Sequencing Project. The program calculates the log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leismania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.
CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg) which is preceded by a stretch of pyrimidines or part of a Kozak sequence. If this cannot be identified we choose the most upstream initiation codon.
Transmembrane domains were predicted as implemented at the TMHMM server: http://www.cbs.dtu.dk/services/TMHMM-1.0/
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid L673 is overlapped at the 5' end by cosmid L2581, and at the 3' end (possibly) by L1063.
FEATURES Location/Qualifiers

source 1. .21057
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="23"
/clone="cosmid L673"
gene 1. .4716
/gene="L673.01"
CDS 1. .4716
/gene="L673.01"
/note="L673.01, len = 1570 aa, multidrug resistance protein, copy 1; BLASTP similarity to copy 2 (L673.02), largest region: identities = 1141/1522 (74%), positives = 1265/1522 (82%); contains two matches to PROSITE PS00017 ATP/GTP-binding site motif A (P-loop), two matches to PROSITE PS00211 ABC transporters family signature, a match to PROSITE PS00030 Eukaryotic putative RNA-binding region RNP-1 signature; contains two Pfam matches to entry PF00664 ABC_membrane, ABC transporter transmembrane region, two Pfam matches to entry PF00005 ABC_tran, ABC transporter; predicted TM helices at aa 243-265, 357-379, 384-406, 473-495, 962-984, 997-1015, 1098-1116, 1189-1207; coiled-coil region
aa 870-885; good similarity to MDR_LEITA, multidrug resistance protein (1548 aa, Leishmania tarentolae, EMBL: X17154, CAA35038); Fasta scores: E():0, 51.0% identity in 1530 aa"
/codon_start=1
/label=L673.01
/product="multidrug resistance protein, copy 1"
/protein_id="CAB64568.1"
/db_xref="GI:6635093"
/translation="MLQETIDEELSMEMTEALPCAPPPGRSRHQSDPLYDYRSEHLAL
QARLAEIWGPEAPYTPVPEETSSWFKRYSYGWVYETVVLASKEKLTHEALPPPTRDVR
AHECGLRLSRAVQAAMYERNAWNCMVGTEVVSTLDASSRGVLRWVGVPQQGGYTRMMA
GVEWSVPPALRTAARSDDSGASPFFDGVAHGEHLFTPEQSGMSTLEETTTLRLDLTSR
GGVVEIPTPKRLPLTRLLAFKLPYYLWMQVPFLLVSNICVITLPSILQAFVAFLDDDP
GTQTWGRGLGLVAGIFLVQAMQSVCNQRFNYISFRCGLQYRSALNSLIFEKCMIISSK
SLAQPDMNAGRVINMMSTDVERVYFFMLFCMFLWSSPVVLLLAIVQLWRLVGWCAIIA
VASFLATIPLNAFFMGIQMRARRDIAKAADARVKATSEFFSGIRVAKFMTWEPCFVAN
IEEKRAVELFFLNKIQNARVATSFVNNATPMLMIALVFTVYYCTGHELSTTVVFPTIA
LLGVLRQPFQQIPWVFTIAVQFLISIGRIRRFLECDDATCSTVQDMEEYWREQREHSA
ACQLAAVLEKVDVTAFVPVKLPFAPKVKTSLLSRALRMLCCEQCKPTKRHPPPSVVVE
NTGYVSPASASRHIVEGCSGTVHTATPTSTRSAKAPQMKTDEFFELEPKVLLRDVSVS
VLRGKLTVVLGATGSGKSTLLQSLLSQFEISEGRVWAERSIAYVPQQAWIMNATVRSN
ILFFDEEDAARLADAVRVSQLEADLAQLSGGLETEIGEKGVNLSGGQKARVSLARAVY
ANRDVYLLDDPLSALDAHVGERVVEECFLGALAGKTRVLATHQVHIVPRADYVVALGD
ERVEFSGSSADFMRTSIYAGMASGLTVNKEEGDAESSESALDAEEEEAEVMVDRSASK
GEGPDEEDDAAARDPAARSFVVEEEKASGGVPWRTYLAYFRYCGGAHVAATIVLVFAV
TELITVSSSVWLSRWTTKEEGESNALYLTVYLLMVLVGGAGYPFRFVVSYTAMRHGCS
ALHRTILRSVTAGTIEFFDRTPLGRLLNRFSRDIDTLDNGLQMSTISLLECLFSISVS
MLVTTYSQPFVLIALAPCGYLYYRLMVFYNSANREIRRYTSVMKTPVFTQLTELTNGL
ATITAYGKAKKVMAEALERLDVVHSCGNLENNTNRWLAVRVELLSNIVVTAIAFVGVI
TTCLRSNRIDIGLISLSLTMAMQTTGELNWLVRMVATMEADMNSVERLQYYIDHVPKE
AMPELDAEVDALERRTGMAADVTGTVVIEPASPTGAAPHIVQAGSLVFEGVQMRYREG
LPLVLRGVSFRIAPREKVGIVGRTGSGKSTLLLTFMRMVEVCGGEIRVNGREIGSYGL
RELRRQFSMIPQDPVLFDGTVRQNVDPFLEASSAEVWAALELVGLRERVASESEGIDS
CVLEGGSNYSVGQRQLMCMARALLKKGSGFILMDEATANIDPALDRQIQATVMSAFSA
YTVITIAHRLHTVAQYDKIIVMDHGAVAEMGSPRELVMNRQSIFHSMVEAGGPSARRH
FLSLVRRRDNESAPYSKCG"
misc_feature complement(1. .2136)
/note="terminal region of BLASTN similarity to bases 1. .2136 AL133436 Leishmania major Friedlin chromosome 23 cosmid L7276, 100% identity over 2135 bases"
misc_feature 118. .4549
/gene="L673.01"
/note="general region of BLASTN similarity to: X17154 Leishmania tarentolae H circle borne ltpgpA gene for P-glycoprotein, L29484 Leishmania tarentolae P-glycoprotein related protein, L29485 Leishmania tarentolae P-glycoprotein related protein, U55381 Leishmania tropica P-glycoprotein E gene, U63320 Leishmania tropica P-glycoprotein (ltrmdr1) gene, U95956

Trypanosoma cruzi P-glycoprotein (tcpgp1A) gene, Z49222 T.cruzi gene for P-glycoprotein, U95956 Trypanosoma cruzi P-glycoprotein (tcpgp1A) gene, U89026 Trypanosoma brucei brucei putative ABC transporter 1 (Tbabc1) gene, AI676508 etmEST0269 EtH1 Eimeria tenella cDNA clone etmf004 5', mRNA sequence"
repeat_region 283. .4631
/note="repeated region, see bases 8444. .12798 of this cosmid sequence; three blocks of high similarity: bases 1-339 (of this feature): 99%, bases 1336-2342: 99%, bases 3371-4349: 97%"
misc_feature 727. .795
/gene="L673.01"
/note="predicted TM helix region"
misc_feature 730. .1551
/gene="L673.01"
/note="Pfam match to entry PF00664 ABC_membrane, ABC transporter transmembrane region., score 89.60, E-value 6.4e-23"
misc_feature 1069. .1137
/gene="L673.01"
/note="predicted TM helix region"
misc_feature 1150. .1218
/gene="L673.01"
/note="predicted TM helix region"
misc_feature 1417. .1485
/gene="L673.01"
/note="predicted TM helix region"
misc_feature 2056. .2571
/gene="L673.01"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 110.80, E-value 2.5e-29"
misc_feature 2077. .2100
/gene="L673.01"
/note="PROSITE PS00017 ATP/GTP-binding site motif A (P-loop), info count = 13.8"
misc_feature 2347. .2391
/gene="L673.01"
/note="PROSITE PS00211 ABC transporters family signature, info count = 19.5"
misc_feature 2860. .3696
/gene="L673.01"
/note="Pfam match to entry PF00664 ABC_membrane, ABC transporter transmembrane region., score 81.40, E-value 1.9e-20"
misc_feature 2884. .2952
/gene="L673.01"
/note="predicted TM helix region"
misc_feature 2989. .3045
/gene="L673.01"
/note="predicted TM helix region"
misc_feature 3292. .3348
/gene="L673.01"

Query Match 75.8%; Score 18.2; DB 3; Length 21057;
Best Local Similarity 87.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 ccgttcgggcgcgaggaagcccg 23
          |||||||| | ||||||||||| |
Db   6567 CCGTTCGGCCACGAGGAAGCCGG 6589
```

RESULT 14
AC004235
LOCUS AC004235 33311 bp DNA linear PRI 27-FEB-1998
DEFINITION Homo sapiens chromosome 16, cosmid clone 432A1 (LANL), complete sequence.
ACCESSION AC004235
VERSION AC004235.1 GI:2914670
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 33311)
AUTHORS Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Misra,M. and Deaven,L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 33311)
AUTHORS Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 33311)
AUTHORS Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Misra,M. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-1998) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
FEATURES Location/Qualifiers
source 1. .33311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="432A1"
repeat_region 2. .361
/rpt_family="Alu"
repeat_region 468. .793
/rpt_family="Alu"
repeat_region complement(793. .999)
/rpt_family="MER29"
repeat_region 1002. .1574
/rpt_family="Alu"
repeat_region 1666. .1975
/rpt_family="Alu"
repeat_region complement(1947. .2237)
/rpt_family="MER7"
repeat_region complement(2267. .2577)
/rpt_family="Alu"
repeat_region 2572. .3411
/rpt_family="L1"
repeat_region 3403. .3736
/rpt_family="Alu"
repeat_region 3759. .4016
/rpt_family="Alu"
repeat_region 4216. .4521
/rpt_family="Alu"
misc_feature 4638. .4677
/note="GRAIL 2 excellent exon, frame 0"
misc_feature 5010. .5421
/note="99% identity zu56e12.r1"
/db_xref="dbEST:AA405422"
STS 5401. .5762
/standard_name="WI-16189"
/note="97% identity"
/db_xref="dbSTS:G24496"
repeat_region complement(7741. .8075)
/rpt_family="Alu"
repeat_region complement(8181. .8498)
/rpt_family="Alu"
repeat_region complement(8974. .9586)
/rpt_family="Alu"
misc_feature 10139. .10226
/note="90% identity nf76b05.s1"
/db_xref="dbEST:Z77852"
misc_feature complement(11820. .11866)

```
                     /note="GRAIL 2 good exon, frame 0"
     repeat_region   complement(12940. .13283)
                     /rpt_family="AluSx"
     repeat_region   13557. .13856
                     /rpt_family="Alu"
     repeat_region   13856. .15120
                     /rpt_family="Alu"
     repeat_region   15532. .15656
                     /rpt_family="Alu"
     repeat_region   complement(15653. .15955)
                     /rpt_family="Alu"
     repeat_region   complement(16004. .16333)
                     /rpt_family="Alu"
     repeat_region   16915. .17497
                     /rpt_family="Alu"
     repeat_region   17283. .17318
                     /note="(AAGG)9"
                     /rpt_type=tandem
                     /rpt_unit=AAGG
     repeat_region   17316. .17339
                     /note="(AGGT)6"
                     /rpt_type=tandem
                     /rpt_unit=AGGT
     repeat_region   18368. .18387
                     /note="(T)20"
                     /rpt_type=tandem
                     /rpt_unit=T
     misc_feature    20809. .20983
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   21436. .21736
                     /rpt_family="Alu"
     misc_feature    22090. .22181
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    22592. .22745
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    22880. .23192
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    23270. .23349
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    23462. .23554
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    complement(23801. .24330)
                     /note="94% identity nd88d11.s1"
                     /db_xref="dbEST:AA581386"
     repeat_region   23966. .23997
                     /note="(A)32"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   24159. .24178
                     /note="(A)20"
                     /rpt_type=tandem
                     /rpt_unit=A
     misc_feature    25593. .25790
                     /note="GRAIL 2 excellent exon, frame 1"
     repeat_region   26821. .26896
                     /rpt_family="MIR"
     misc_feature    27107. .27328
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    27256. .27331
                     /note="88% identity mouse vg33c04.r1"
                     /db_xref="dbEST:AA510585"
     misc_feature    27465. .27884
                     /note="GRAIL 2 good exon, frame 1"
     misc_feature    27674. .27991
                     /note="83% identity mouse vf86g11.r1"
                     /db_xref="dbEST:AA451052"
     misc_feature    27901. .28216
                     /note="97% identity EST15204"
                     /db_xref="dbEST:AA302093"
     gene            complement(join(28743. .29014,29088. .29165,29950. .30107,
                     30195. .30367,30500. .30606,31269. .31762,32614. .>32981))
                     /gene="Myt1"
                     /note="kinase Myt1; U56816"
                     /evidence=experimental
     CDS             complement(join(28903. .29014,29088. .29165,29950. .30107,
                     30195. .30367,30500. .30606,31269. .31762,32614. .>32981))
                     /gene="Myt1"
                     /note="kinase Myt1"
                     /codon_start=3
                     /evidence=experimental
                     /product="Myt1"
                     /protein_id="AAC04477.1"
                     /db_xref="GI:2914671"
                     /translation="PPALAMPMPTEGTPPPLSGTPIPVPAYFRHAEPGFSLKRPRGLS
                     RSLPPPPPAKGSIPISRLFPPRTPGWHQLQPRRVSFRGEASETLQSPGYDPSRPESFF
                     QQSFQRLSRLGHGSYGEVFKVRSKEDGRLYAVKRSMSPFRGPKDRARKLAEVGSHEKV
                     GQHPCCVRLEQAWEEGGILYLQTELCGPSLQQHCEAWGASLPEAQVWGYLRDTLLALA
                     HLHSQGLVHLDVKPANIFLGPRGRCKLGDFGLLVELGTAGAGEVQEGDPRYMAPELLQ
                     GSYGTAADVFSLGLTILEVACNMELPHGGEGWQQLRQGYLPPEFTAGLSSELRSVLVM
                     MLEPDPKLRATAEALLALPVLRQPRAWGVLWCMAAEALSRGWALWQALLALLCWLWHG
                     LAHPASWLQPLGPPATPPGSPPCSLLLDSSLSSNWDDDSLGPSLSPEAVLARTVGSTS
                     TPRSRCTPRDALDLSDINSEPPRGSFPSFEPRNLLSLFEDTLDPT"
     misc_feature    complement(28903. .29014)
                     /gene="Myt1"
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    complement(30195. .30367)
                     /gene="Myt1"
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    complement(30500. .30606)
                     /gene="Myt1"
                     /note="GRAIL 2 excellent exon, frame 0"
     misc_feature    complement(31269. .31762)
                     /gene="Myt1"
                     /note="GRAIL 2 excellent exon, frame 1"
     misc_feature    31702. .31769
                     /note="82% identity mouse vh11f12.r1"
                     /db_xref="dbEST:AA475070"
     repeat_region   31856. .31876
                     /note="(T)21"
                     /rpt_type=tandem
                     /rpt_unit=T
     misc_feature    32609. .32924
                     /note="84% identity mouse vh11f12.r1"
                     /db_xref="dbEST:AA475070"
     misc_feature    complement(32614. .32774)
                     /gene="Myt1"
                     /note="GRAIL 2 excellent exon, frame 2"
BASE COUNT     7028 a    9848 c    9949 g    6486 t
ORIGIN

  Query Match          75.8%;  Score 18.2;  DB 9;  Length 33311;
  Best Local Similarity   87.0%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cgttcgggcgcgaggaagcccgg 24
           |||  |||||||||||| ||||||
Db   20745 CGTAGGGGCGCGAGGATGCCCGG 20767


RESULT  15
AC004396
LOCUS       AC004396               43903 bp    DNA     linear   HTG 13-MAR-2000
DEFINITION  Pseudomonas sp. chromosome Genomic clone cos3b, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC004396
VERSION     AC004396.12  GI:7230801
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Pseudomonas sp.
  ORGANISM  Pseudomonas sp.
            Bacteria; Proteobacteria.
REFERENCE   1  (bases 1 to 43903)
  AUTHORS   Roe,B.A.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43903)
  AUTHORS   Roe,B.A.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:7229744.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15232: contig of 15232 bp in length
            *    15233    15352: gap of unknown length
            *    15353    43903: contig of 28551 bp in length.
FEATURES             Location/Qualifiers
     source          1. .43903
                     /organism="Pseudomonas sp."
                     /db_xref="taxon:306"
                     /chromosome="Genomic"
                     /clone="cos3b"
BASE COUNT     8111 a  13520 c  13821 g   8330 t    121 others
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 2;  Length 43903;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cgttcgggcgcgaggaagcccgg 24
           || || ||||| ||||||||||
Db   37376 CGCTCCGGCGCCAGGAAGCCCGG 37398


Search completed: June 23, 2002, 06:32:15
Job time: 63867 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 15:00:17 ; Search time 652.13 Seconds
(without alignments)
63.187 Million cell updates/sec

Title: US-09-747-514A-3
Perfect score: 24
Sequence: 1 ccgttcgggcgcgaggaagcccgg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18.8	78.3	1241	22	AAI87016	Human polynucleoti
c	2	18.2	75.8	1413	21	AAC55817	S. lavendulae MmcC
c	3	18.2	75.8	53500	21	AAC55842	Complete nucleotid
c	4	17.4	72.5	1496	15	AAQ78165	UL49 homologue gen
c	5	17.4	72.5	1499	21	AAA59128	Nucleotide sequenc
c	6	17.4	72.5	1499	24	AAS18163	DNA encoding bovin
c	7	17.4	72.5	12749	20	AAX78032	Human Huntington's
	8	17.2	71.7	48275	21	AAA81501	N. meningitidis pa
c	9	17.2	71.7	349980	21	AAF21610	Neisseria meningit
	10	16.8	70.0	9431	13	AAQ20907	DNA encoding the c
c	11	16.6	69.2	85	22	AAH29361	Drosophila melanog
	12	16.6	69.2	418	22	AAH46964	Human secreted pro
	13	16.6	69.2	903	20	AAX36806	Human transmembran
c	14	16.6	69.2	1190	22	AAS44946	cDNA encoding nove
	15	16.6	69.2	1237	22	AAI58532	Human polynucleoti
	16	16.6	69.2	1241	12	AAQ11297	Genomic DNA encodi
	17	16.6	69.2	1256	20	AAX36807	Human transmembran
	18	16.6	69.2	1474	22	AAS41367	cDNA encoding nove
	19	16.6	69.2	1658	21	AAF16097	Human prostate can
	20	16.6	69.2	1658	22	AAH33275	Human colon cancer
	21	16.6	69.2	1773	23	AAS51462	Pseudomonas aerugi
	22	16.6	69.2	1811	22	AAH23805	Human transferase
	23	16.6	69.2	2888	22	AAI60318	Human polynucleoti
c	24	16.6	69.2	34094	20	AAZ30163	Complete nucleotid
	25	16.2	67.5	390	18	AAT59695	2,4-Diacetylphloro
	26	16.2	67.5	606	18	AAT59693	2,4-Diacetylphloro
c	27	16.2	67.5	1878	21	AAF15938	Human prostate can
c	28	16.2	67.5	2629	23	ABL05033	Drosophila melanog
c	29	16.2	67.5	4629	23	ABL05032	Drosophila melanog
c	30	16.2	67.5	4979	23	AAS84830	DNA encoding novel
	31	16.2	67.5	6170	18	AAT59687	Genes for (modulat
	32	16.2	67.5	6387	18	AAT59686	Genes for (modulat
	33	16.2	67.5	7106	14	AAQ47193	Plasmid pHCV-167 c
	34	16.2	67.5	7106	16	AAQ97494	pHCV167 sequence.
	35	16.2	67.5	7198	20	AAX60278	DNA encoding a 2,4
	36	16.2	67.5	7298	14	AAQ47192	Plasmid pHCV-162 c
	37	16.2	67.5	29879	14	AAQ46806	eryA region of S.
	38	16.2	67.5	32572	24	AAS17820	PAdV-5 HNF-70 comp
c	39	16.2	67.5	4403765	22	AAI99683	Mycobacterium tube
c	40	16.2	67.5	4411529	22	AAI99682	Mycobacterium tube
c	41	16	66.7	792	20	AAZ42110	Human endometrium
c	42	16	66.7	894	23	AAS51512	Pseudomonas aerugi
c	43	16	66.7	928	22	AAK91820	Human cDNA 5'-end
c	44	16	66.7	928	22	AAK93219	Human cDNA clone r
	45	16	66.7	993	22	AAH66060	C glutamicum codin

ALIGNMENTS

RESULT 1
AAI87016
ID AAI87016 standard; cDNA; 1241 BP.
XX
AC AAI87016;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 7076.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO07085.

XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 7076; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1241 BP; 293 A; 301 C; 341 G; 292 T; 14 other;

Query Match 78.3%; Score 18.8; DB 22; Length 1241;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 gttcgggcgcgaggaagcccgg 24
|| |||||||||||||| ||||||
Db 844 gctcgggcgcgaggaggcccgg 865

RESULT 2
AAC55817/c
ID AAC55817 standard; DNA; 1413 BP.
XX
AC AAC55817;
XX
DT 19-JAN-2001 (first entry)
XX
DE S. lavendulae MmcC encoding DNA sequence.
XX
KW Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX
PN WO200053737-A2.
XX
PD 14-SEP-2000.
XX
PF 10-MAR-2000; 2000WO-US06394.
XX
PR 12-MAR-1999; 99US-0266965.
XX
PA (MINU) UNIV MINNESOTA.
PA (SHER/) SHERMAN D H.
PA (MAOY/) MAO Y.
PA (VARO/) VAROGLU M.
PA (HEMM/) HE M.
PA (SHEL/) SHELDON P C.
XX
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
DR WPI; 2000-601980/57.
DR P-PSDB; AAB32516.
XX
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT the molecular basis of mitosane ring system biosynthesis -
XX
PS Disclosure; Page 271-272; 399pp; English.
XX
CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzoquinone and axiridine ring systems. The S. lavendulae
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC spanning 55kb of DNA. The invention includes an expression cassette
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC and host cells transformed with the cassette. The nucleotide, and protein
CC sequences and the transformed host cells of the invention result in
CC antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC antibiotic activities. The nucleotide sequences are used to elucidate the
CC molecular basis for the biosynthesis of the mitosane ring system, as well
CC as to engineer the biosynthesis of novel natural products, e.g.
CC antibiotics, anti-inflammatory agents, anti-cancer agents,
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC obstructive pulmonary disease as well as other disease involving
CC respiratory inflammation, or cholesterol-lowering agents or as crop
CC protection agents (e.g. fungicides or insecticides) as well as
CC bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC sequences and encoded proteins. Sequences AAC55812-C55814,
CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC cloning of the mitomycin biosynthetic genes.
XX
SQ Sequence 1413 BP; 180 A; 609 C; 447 G; 177 T; 0 other;

Query Match 75.8%; Score 18.2; DB 21; Length 1413;
Best Local Similarity 87.0%; Pred. No. 62;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgttcgggcgcgaggaagcccg 23
| |||||||||||| |||||| |
Db 399 CAGTTCGGGCGCGACGAAGCCGG 377

RESULT 3
AAC55842/c
ID AAC55842 standard; DNA; 53500 BP.
XX
AC AAC55842;
XX
DT 19-JAN-2001 (first entry)
XX
DE Complete nucleotide sequence of the mitomycin biosynthetic genes.
XX
KW Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX
PN WO200053737-A2.
XX
PD 14-SEP-2000.
XX
PF 10-MAR-2000; 2000WO-US06394.
XX
PR 12-MAR-1999; 99US-0266965.
XX
PA (MINU) UNIV MINNESOTA.
PA (SHER/) SHERMAN D H.
PA (MAOY/) MAO Y.
PA (VARO/) VAROGLU M.
PA (HEMM/) HE M.
PA (SHEL/) SHELDON P C.
XX

PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
DR WPI; 2000-601980/57.
XX
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT the molecular basis of mitosane ring system biosynthesis -
XX
PS Disclosure; Figure 26; 399pp; English.
XX
CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzoquinone and axiridine ring systems. The S. lavendulae
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC spanning 55kb of DNA. The invention includes an expression cassette
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC and host cells transformed with the cassette. The nucleotide, and protein
CC sequences and the transformed host cells of the invention result in
CC antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC antibiotic activities. The nucleotide sequences are used to elucidate the
CC molecular basis for the biosynthesis of the mitosane ring system, as well
CC as to engineer the biosynthesis of novel natural products, e.g.
CC antibiotics, anti-inflammatory agents, anti-cancer agents,
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC obstructive pulmonary disease as well as other disease involving
CC respiratory inflammation, or cholesterol-lowering agents or as crop
CC protection agents (e.g. fungicides or insecticides) as well as
CC bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC sequences and encoded proteins. Sequences AAC55812-C55814,
CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC cloning of the mitomycin biosynthetic genes.
XX
SQ Sequence 53500 BP; 7481 A; 19740 C; 19126 G; 7153 T; 0 other;

Query Match 75.8%; Score 18.2; DB 21; Length 53500;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccgttcgggcgcgaggaagcccg 23
|| |||||||||||| |||||| |
Db 24533 CAGTTCGGGCGCGACGAAGCCGG 24511

RESULT 4
AAQ78165/c
ID AAQ78165 standard; DNA; 1496 BP.
XX
AC AAQ78165;
XX
DT 07-JUN-1995 (first entry)
XX
DE UL49 homologue gene of bovine herpes virus 1 (BHV-1).
XX
KW Cattle pathogen; infectious bovine rhinotracheitis; nucleocapsid;
KW VP4; promoter; ss.
XX
OS Bovine herpes virus 1.
XX
FH Key Location/Qualifiers
FT CDS complement (28..26)
FT /*tag= a
FT /label= dUTPase (UL50) ORF start codon
FT CDS 1..266
FT /*tag= b
FT /label= UL49.5 gene
FT GC_signal 256..258
FT /*tag= c
FT TATA_signal 294..300
FT /*tag= d
FT RBS 333..335
FT /*tag= e
FT polyA_signal 1226..1231
FT /*tag= f
FT CDS 1351..1496
FT /*tag= g
FT /label= UL48 ORF
XX
PN WO9424296-A.
XX
PD 27-OCT-1994.
XX
PF 19-APR-1994; 94WO-CA00201.
XX
PR 19-APR-1993; 93US-0051448.
XX
PA (UYSA-) UNIV SASKATCHEWAN.
XX
PI Liang X, Rabuik LA, Zamb T;
XX
DR WPI; 1994-341875/42.
DR P-PSDB; AAR63461.
XX
PT Mutant bovine herpes type 1 virus with deleted gene(s) - used as
PT vaccine, also with replacement bacterial or viral pathogen
PT immunogens and/or cytokine(s), inserted at deletion
XX
PS Example; Fig 11; 109pp; English.
XX
CC AAQ78165/R63461 are the nucleotide sequence and deduced AA sequence
CC of BHV 1 UL49 homologue gene. The complete DNA sequence between the
CC UL48 homologue gene and dUTPase gene (UL50 homologue) gene are
CC presented. Putative G-C box, TATA box and polyA signal sequence
CC are indicated in FT. The boundaries of UL50, UL 49.5 and UL48
CC ORFs surrounding the UL 49 homologue gene are also indicated.
XX
SQ Sequence 1496 BP; 212 A; 535 C; 517 G; 232 T; 0 other;

Query Match 72.5%; Score 17.4; DB 15; Length 1496;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 cgggcgcgaggaagcccgg 24
|||||||||||| ||||||
Db 695 CGGGCGCGAGGACGCCCGG 677

RESULT 5
AAA59128/c
ID AAA59128 standard; DNA; 1499 BP.
XX
AC AAA59128;
XX
DT 07-NOV-2000 (first entry)
XX
DE Nucleotide sequence of the UL49 gene homologue of BHV-1.
XX
KW UL49; homologue; Bovine herpesvirus 1; BHV-1; attenuated virus;
KW essential gene; BHV-1 infection; vaccine; cattle; ss.
XX
OS Bovine herpesvirus.
XX
FH Key Location/Qualifiers
FT CDS 390..1166
FT /*tag= a
FT /product= "UL49 homologue"
XX
PN US6086902-A.
XX
PD 11-JUL-2000.
XX
PF 09-SEP-1994; 94US-0303861.

```
XX
PR   19-APR-1993;   93US-0051448.
XX
PA   (UYSA-) UNIV SASKATCHEWAN.
XX
PI   Babiuk LA,  Zamb T,  Liang X;
XX
DR   WPI; 2000-531327/48.
DR   P-PSDB; AAB07662.
XX
PT   Immunogenic composition useful as vaccine against herpesvirus infection
PT   comprises a live attenuated bovine herpesvirus 1 comprising a mutation
PT   in at least one essential gene to reduce its virulence -
XX
PS   Example 15; Fig 11; 56pp; English.
XX
CC   The present sequence encodes a Bovine herpesvirus 1 (BHV-1) UL49
CC   gene homologue polypeptide. The gene was altered to create the viruses
CC   of the invention. The specification describes an immunogenic composition
CC   comprising a live attenuated BHV-1, which has a mutation in at
CC   least one essential gene of wild type BHV-1 to reduce the
CC   virulence of the virus. The immunogenic composition is useful for
CC   treating or preventing BHV-1 infection in a bovine host. It is also
CC   useful as vaccine to immunize cattle against infection with wild
CC   type BHV-1.
XX
SQ   Sequence 1499 BP; 212 A; 536 C; 519 G; 232 T; 0 other;


  Query Match             72.5%;  Score 17.4;  DB 21;  Length 1499;
  Best Local Similarity   94.7%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6 cgggcgcgaggaagcccgg 24
          ||||||||||||| ||||||
Db    695 CGGGCGCGAGGACGCCCGG 677


RESULT   6
AAS18163/c
ID   AAS18163 standard; DNA; 1499 BP.
XX
AC   AAS18163;
XX
DT   26-MAR-2002  (first entry)
XX
DE   DNA encoding bovine herpesvirus protein.
XX
KW   Bovine herpesvirus protein; gene therapy; reporter protein; ds.
XX
OS   Bovine herpesvirus 1.
XX
FH   Key             Location/Qualifiers
FT   CDS             390..1166
FT                   /*tag=  a
FT                   /product= "Bovine herpesvirus protein"
XX
PN   US6316252-B1.
XX
PD   13-NOV-2001.
XX
PF   17-DEC-1998;   98US-0213343.
XX
PR   17-DEC-1998;   98US-0213343.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Harms JS,  Splitter GA;
XX
DR   WPI; 2002-074374/10.
DR   P-PSDB; AAU11367.
XX
PT   New bovine herpes virus protein linked to a non-bovine reporter
PT   protein, useful for delivering therapeutic and/or reporter proteins to
PT   mammalian cells, either in vitro or in vivo -
XX
PS   Example 2; Column 5-8; 8pp; English.
XX
CC   The invention relates to a peptide comprising bovine herpesvirus protein
CC   linked to a non-bovine reporter protein. The peptide is useful for
CC   delivering therapeutic and/or reporter proteins to mammalian cells,
CC   either in vitro or in vivo. It is especially suited to deliver proteins
CC   to ruminant and primate cells, thus has in vivo and in vitro
CC   pharmaceutical, veterinary and research applications. The system permits
CC   delivery of effector proteins deep within tissue. This sequence
CC   represents DNA encoding the bovine herpesvirus protein of the invention.
XX
SQ   Sequence 1499 BP; 212 A; 536 C; 519 G; 232 T; 0 other;


  Query Match             72.5%;  Score 17.4;  DB 24;  Length 1499;
  Best Local Similarity   94.7%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6 cgggcgcgaggaagcccgg 24
          ||||||||||||| ||||||
Db    695 CGGGCGCGAGGACGCCCGG 677


RESULT   7
AAX78032/c
ID   AAX78032 standard; cDNA to mRNA; 12749 BP.
XX
AC   AAX78032;
XX
DT   19-AUG-1999  (first entry)
XX
DE   Human Huntington's chorea associated protein homologue cDNA.
XX
KW   Huntington's chorea; human; model; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..9547
FT                   /*tag=  a
FT                   /product= "Huntington's chorea associated protein
FT                            homologue"
XX
PN   JP11137258-A.
XX
PD   25-MAY-1999.
XX
PF   14-NOV-1997;   97JP-0314020.
XX
PR   14-NOV-1997;   97JP-0314020.
XX
PA   (SLAK-) SLA KENKYUSHO KK.
XX
DR   WPI; 1999-374378/32.
DR   P-PSDB; AAY08898.
XX
PT   cDNA of huntington's gene - useful for development of animal
PT   model of Huntington's disease
XX
PS   Claim 1; Page 6-10; 18pp; Japanese.
XX
CC   This invention describes a novel human Huntington's chorea associated
CC   protein homologue. The cDNA and the other material of the invention
CC   are useful for the development of a model animal of Huntington's disease.
XX
SQ   Sequence 12749 BP; 2292 A; 4064 C; 3925 G; 2468 T; 0 other;


  Query Match             72.5%;  Score 17.4;  DB 20;  Length 12749;
  Best Local Similarity   94.7%;  Pred. No. 1.3e+02;
```

```
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 cgggcgcgaggaagcccgg 24
           |||||||||||||| ||||
Db    7292 CGGGCGCGAGGAAGGCCGG 7274


RESULT   8
AAA81501
ID   AAA81501 standard; DNA; 48275 BP.
XX
AC   AAA81501;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_48 SEQ ID NO:48.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 1317-1331; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 48275 BP; 10709 A; 12099 C; 13497 G; 11970 T; 0 other;

  Query Match             71.7%;  Score 17.2;  DB 21;  Length 48275;
  Best Local Similarity   86.4%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cgttcgggcgcgaggaagcccg 23
           ||||||||| |||||||||| |
Db   27591 cgttcgggcaggaggaagccgg 27612


RESULT   9
AAF21610/c
ID   AAF21610 standard; DNA; 349980 BP.
XX
AC   AAF21610;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
```

```
XX
SQ   Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;


  Query Match             71.7%;  Score 17.2;  DB 21;  Length 349980;
  Best Local Similarity   86.4%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cgttcgggcgcgaggaagcccg 23
          |||||||||  ||||||||| |
Db 297563 CGTTCGGGCAGGAGGAAGCCGG 297542


RESULT  10
AAQ20907
ID   AAQ20907 standard; DNA; 9431 BP.
XX
AC   AAQ20907;
XX
DT   22-MAY-1992  (first entry)
XX
DE   DNA encoding the chimeric protein V1V2-hCH2-KA.
XX
KW   CD4; LTI; Streptomyces longisporous; HIV gp120; AIDS; IgG1; T cell;
KW   immunoglobulin; surface glycoprotein; virus; MHC class II; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   signal_peptide  648..731
FT                   /*tag= a
FT   mat_peptide     732..1286
FT                   /*tag= b
FT                   /product= V1V2
FT   misc_feature    1287..1331
FT                   /*tag= c
FT                   /function= hinge_region
FT   mat_peptide     1332..1680
FT                   /*tag= d
FT                   /product= CH2
XX
PN   WO9200985-A.
XX
PD   23-JAN-1992.
XX
PF   01-JUL-1991;   91WO-US04663.
XX
PR   05-MAR-1991;   91US-0665218.
PR   11-JUL-1990;   90US-0551584.
XX
PA   (SMIK ) SMITHKLINE BEECHAM.
XX
PI   Brawner ME,  Fornwald JA,  Arthos J;
XX
DR   WPI; 1992-056814/07.
DR   P-PSDB; AAR20634.
XX
PT   Nucleic acid sequences for production of CD4 chimeric protein -
PT   used to transfect streptomyces, contg. LTI signal sequence linked
PT   to pro-peptide sequence facilitating peptide cleavage
XX
PS   Claim 10; Page 23; 47pp; English.
XX
CC   The sequence was obtd. by sequencing the plasmid vetor V1V2-hCH2-KA.
CC   Plasmid V1V2-hCH2strept was transformed into S. lividans strain 1326
CC   and transformants selected. V1V2-hCH2-KA was expressed at 14mg/l.
CC   The plasmid construct contains a CD4 chimera (V1V2) in which the
CC   carboxy terminal portion of the encoded protein consists of a murine
CC   immunoglobulin light chain constant region.  This sequence is
CC   operably linked to the coding sequence of the signal peptide of
CC   Streptomyces LTI, modified at its 5' end by the addition of bases
CC   encoding Lys-Arg. Also included in the expression vector is the
CC   sequence encoding an IgG1 constant region comprising the hinge and
CC   CH2 motifs.  Human IgG1 is the most effective immunoglobulin
CC   subclass at mediating cell killing by both complement and ADCC.
CC   The vectors are used for the prodn. of sol. CD4 chimeric proteins
CC   in bacterial hosts,in which the HIV gp120 binding region is joined
CC   to a region of the human Ig constant region lacking the CH3 domain,
CC   which increases the stability of the CD4, thus increasing the serum
CC   half life and/or potency against HIV infection and inhibit
CC   virus-induced cell fusion, relative to soluble CD4.
CC   See also AAQ20908,9.
XX
SQ   Sequence 9431 BP; 1712 A; 2938 C; 2964 G; 1808 T; 9 other;


  Query Match             70.0%;  Score 16.8;  DB 13;  Length 9431;
  Best Local Similarity   90.0%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 cgttcgggcgcgaggaagcc 21
          ||| || ||||||||||||||
Db   4836 cgtgcgcgcgcgaggaagcc 4855


RESULT  11
AAH29361/c
ID   AAH29361 standard; DNA; 85 BP.
XX
AC   AAH29361;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Drosophila melanogaster essential gene fragment, SEQ ID NO: 550.
XX
KW   Drosophila melanogaster; fruit fly; essential gene; screening assay;
KW   pesticide; crop protection; chromosome 3; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200118547-A1.
XX
PD   15-MAR-2001.
XX
PF   06-SEP-2000; 2000WO-GB03444.
XX
PR   07-SEP-1999;   99GB-0021009.
XX
PA   (UNIU ) UNIV GLASGOW.
XX
PI   Davies RW,  Kaiser K,  Yang MY;
XX
DR   WPI; 2001-281436/29.
XX
PT   Screening assays for used for identifying compounds having a
PT   physiological effect on proteins identified as being essential -
XX
PS   Claim 23; Page 483; 695pp; English.
XX
CC   The present sequence is part of an essential gene from Drosophila
CC   melanogaster. Lack of expression of the protein encoded by this
CC   gene leads to a lethal or semi-lethal phenotype. The invention
CC   relates to 902 nucleic acid sequences from genes encoding proteins
CC   which are thought to be essential, and to a screening assay for
CC   identifying compounds which have a physiological effect on these
CC   proteins. Suitable compounds are useful as pesticides and may be used
CC   in conjunction with other pesticides and herbicides for crop
CC   protection. The gene corresponding to the present sequence is located
CC   on chromosome 3.
XX
SQ   Sequence 85 BP; 27 A; 19 C; 24 G; 15 T; 0 other;


  Query Match             69.2%;  Score 16.6;  DB 22;  Length 85;
  Best Local Similarity   82.6%;  Pred. No. 3.3e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       2 cgttcgggcgcgaggaagcccgg 24
           | ||||||||||| |||| |||
Db      46 CTTTCGGGCGCGAATAAGCGCGG 24


RESULT  12
AAH46964
ID   AAH46964 standard; cDNA; 418 BP.
XX
AC   AAH46964;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human secreted protein encoding cDNA (clone Id HSLIA81).
XX
KW   Secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   nootropic; neuroprotective; antibacterial; virucide; fungicide; human;
KW   opthalmalogical; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155430-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01431.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   12-SEP-2000; 2000US-0231968.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Komatsoulis GA,  Baker KP,  Birse CE,  Soppet DR,  Olsen HS;
PI   Moore PA,  Wei P,  Ebner R,  Duan DR,  Shi Y,  Choi GH,  Fiscella M;
PI   Ni J,  Ruben SM,  Barash SC;
XX
DR   WPI; 2001-476220/51.
DR   P-PSDB; AAB85554.
XX
PT   17 isolated nucleic acid molecules encoding human secreted proteins,
PT   used to preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 436; 482pp; English.
XX
CC   The invention provides novel human secreted proteins and polynucleotides
CC   encoding them. The secreted proteins can be expressed by standard
CC   recombinant methodology. The secreted proteins and polynucleotides are
CC   used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC   mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
CC   also be used in diagnosing a pathological condition. The antibodies to
CC   the proteins can also be used in alleviating symptoms associated with the
CC   disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
CC   linked immunosorbent assays (ELISA). Disorders which are diagnosed or
CC   treated include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC   ocular disorders e.g. corneal infection. The polypeptides can also be
CC   used to aid wound healing and epithelial cell proliferation, to prevent
CC   skin aging due to sunburn, to maintain organs before transplantation, for
CC   supporting cell culture of primary tissues, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities. The present
CC   sequence represents a human secreted protein encoding cDNA.
XX
SQ   Sequence 418 BP; 66 A; 145 C; 127 G; 72 T; 8 other;

  Query Match             69.2%;  Score 16.6;  DB 22;  Length 418;
  Best Local Similarity   69.6%;  Pred. No. 3.2e+02;
  Matches   16;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cgttcgggcgcgaggaagcccgg 24
           |||:||| | | |:||:|:|||| 
Db     390 cgtycggcctccargamgsccgg 412


RESULT  13
AAX36806
ID   AAX36806 standard; DNA; 903 BP.
XX
AC   AAX36806;
XX
DT   14-JUL-1999  (first entry)
XX
DE   Human transmembrane protein coding sequence, HP01606.
XX
KW   Transmembrane protein; human; cell membrane; proliferation; diagnosis;
KW   cell differentiation; carcinostatic agent; probe; gene therapy;
KW   signal transduction; apoptosis; inhibitor;
KW   phopshatidylethanolamine N-methyltransferase; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918203-A2.
XX
PD   15-APR-1999.
XX
PF   05-OCT-1998;   98WO-JP04475.
XX
PR   08-OCT-1997;   97JP-0276271.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1999-277268/23.
DR   P-PSDB; AAY13941.
XX
PT   Human transmembrane proteins and nucleotide sequences
XX
PS   Claim 3; Page 98-99; 139pp; English.
XX
CC   This sequence encodes a human transmembrane protein of the invention.
CC   All of the proteins exist in the cell membrane, so are considered to be
CC   proteins controlling the proliferation and differentiation of the cells.
CC   They may be useful as carcinostatic agents or as antigens for preparing
CC   antibodies against the proteins. The cDNAs can be used as probes for
CC   gene diagnosis and gene sources for gene therapy, as well as for
CC   large-scale expression of the proteins. The HP01498 (see AAY13939)
CC   protein may be associated with signal transduction associated with
CC   apoptosis, and therefore useful in inhibition of apoptosis. The HP01962
CC   (see AAY13943) protein can be used to treat diseases associated with
CC   phopshatidylethanolamine N-methyltransferase. The proteins are
CC   identified by the presence of a hydrophobic transmembrane region,
CC   knowledge of the protein function is not required, as in e.g. methods of
CC   expression cloning.
XX
SQ   Sequence 903 BP; 190 A; 234 C; 242 G; 237 T; 0 other;

  Query Match             69.2%;  Score 16.6;  DB 20;  Length 903;
  Best Local Similarity   82.6%;  Pred. No. 3.1e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ccgttcgggcgcgaggaagcccg 23
           | |||||||||||| ||||| | |
Db     204 ctgttcgggcgcggggaaggcgg 226


RESULT  14
```

```
AAS44946/c
ID   AAS44946 standard; cDNA; 1190 BP.
XX
AC   AAS44946;
XX
DT   18-DEC-2001  (first entry)
XX
DE   cDNA encoding novel human secretory protein, Seq ID No 27.
XX
KW   Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW   ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW   transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW   amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW   ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW   gut protection; lung; liver fibrosis; immune deficiency; infection;
KW   severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW   multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW   fertility; analgesic; pain; antigen; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166689-A2.
XX
PD   13-SEP-2001.
XX
PF   05-MAR-2001; 2001WO-US04942.
XX
PR   07-MAR-2000; 2000US-0519705.
PR   19-MAY-2000; 2000US-0574454.
PR   17-JUN-2000; 2000US-0596193.
PR   14-JUL-2000; 2000US-0616847.
PR   19-SEP-2000; 2000US-0665363.
PR   20-OCT-2000; 2000US-0693267.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Xu C,  Wehrman T,  Ren F,  Ma Y,  Zhou P;
PI   Zhao QA,  Yang Y,  Drmanac RT,  Zhang J,  Chen R,  Xue AJ,  Wang J;
XX
DR   WPI; 2001-589934/66.
DR   P-PSDB; AAU28046.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis and treatment of
PT   cancer, neurological, inflammatory, and autoimmune disorders -
XX
PS   Claim 1; SEQ ID No 27; 107pp; English.
XX
CC   The invention relates to novel isolated human secreted polypeptides (I)
CC   and polynucleotides (II). (I) and (II) are useful for treating
CC   inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC   ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC   involved in increasing haematopoiesis, stem cell survival, bone growth
CC   and remodeling. (I), (II) and modulators of (II) are useful for
CC   prophylaxis or treatment of one or more cancers. (II) is also useful for
CC   creating transgenic animals useful for studying the in vivo activities of
CC   the polypeptide as well as for studying modulators of the polypeptides.
CC   (I) induces the proliferation of neural cells and regeneration of nerve
CC   and brain tissue and is useful for the treatment of central and
CC   peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC   Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC   sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis and is useful for treating myeloid
CC   or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC   and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC   tissue growth, and in tissue repair, healing of burns, incisions,
CC   ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
CC   disorders, or periodontal disease. Furthermore, (I) is also useful for
CC   gut protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, various immune deficiencies and
CC   disorders including severe combined immunodeficiency (SCID), bacterial or
CC   fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC   rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC   reactions and conditions, such as asthma or other respiratory problems.
CC   In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC   fertility, metabolism, catabolism, anabolism, storage or elimination of
CC   dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC   analgesic effects or other pain reducing effects, immunoglobulin like
CC   activity and can act as an antigen in a vaccine composition to raise an
CC   immune response. AAS44920-AAS45295 represent novel human secreted protein
CC   coding sequences of the invention.
XX
SQ   Sequence 1190 BP; 308 A; 308 C; 309 G; 265 T; 0 other;


  Query Match             69.2%;  Score 16.6;  DB 22;  Length 1190;
  Best Local Similarity   82.6%;  Pred. No. 3.1e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ccgttcgggcgcgaggaagcccg 23
             |||  ||| ||||||||||| |||
Db      46 CCGGCCGGACGCGAGGAAGGCCG 24


RESULT  15
AAI58532
ID   AAI58532 standard; cDNA; 1237 BP.
XX
AC   AAI58532;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 735.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39376.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 735; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
```

CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1237 BP; 302 A; 291 C; 316 G; 328 T; 0 other;

Query Match 69.2%; Score 16.6; DB 22; Length 1237;
Best Local Similarity 82.6%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 ccgttcgggcgcgaggaagcccg 23
           | ||||||||||| ||||| | |
Db     278 ctgttcgggcgcggggaaggcgg 300
```

Search completed: June 23, 2002, 15:01:12
Job time: 81204 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 15:09:02 ; Search time 167.81 Seconds
(without alignments)
35.130 Million cell updates/sec

Title: US-09-747-514A-3
Perfect score: 24
Sequence: 1 ccgttcgggcgcgaggaagcccgg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	17.4	72.5	1499	3	US-08-303-861-17	Sequence 17, Appl
c	2	17.4	72.5	1499	4	US-09-213-343-1	Sequence 1, Appli
	3	16.2	67.5	390	2	US-08-494-907-19	Sequence 19, Appl
	4	16.2	67.5	390	5	PCT-US96-10986-19	Sequence 19, Appl
	5	16.2	67.5	606	2	US-08-494-907-15	Sequence 15, Appl
	6	16.2	67.5	606	5	PCT-US96-10986-15	Sequence 15, Appl
	7	16.2	67.5	6170	2	US-08-494-907-4	Sequence 4, Appli
	8	16.2	67.5	6170	5	PCT-US96-10986-4	Sequence 4, Appli
	9	16.2	67.5	6387	2	US-08-494-907-3	Sequence 3, Appli
	10	16.2	67.5	6387	5	PCT-US96-10986-3	Sequence 3, Appli
	11	16.2	67.5	7106	1	US-08-188-281B-8	Sequence 8, Appli
	12	16.2	67.5	7106	1	US-08-453-552-5	Sequence 5, Appli
	13	16.2	67.5	7106	2	US-08-710-637-5	Sequence 5, Appli
	14	16.2	67.5	7106	5	PCT-US93-00907-5	Sequence 5, Appli
	15	16.2	67.5	7106	5	PCT-US94-07280-8	Sequence 8, Appli
	16	16.2	67.5	7106	5	PCT-US95-01087-8	Sequence 8, Appli
	17	16.2	67.5	7198	4	US-08-994-035C-4	Sequence 4, Appli
	18	16.2	67.5	7298	1	US-08-453-552-3	Sequence 3, Appli
	19	16.2	67.5	7298	2	US-08-710-637-3	Sequence 3, Appli
	20	16.2	67.5	7298	5	PCT-US93-00907-3	Sequence 3, Appli
	21	16.2	67.5	11219	1	US-07-642-734C-1	Sequence 1, Appli
	22	16.2	67.5	11219	3	US-08-439-009A-1	Sequence 1, Appli
c	23	16.2	67.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	24	16.2	67.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	25	15.8	65.8	773	3	US-08-934-131-2	Sequence 2, Appli
c	26	15.8	65.8	8056	3	US-09-136-605-14	Sequence 14, Appl
c	27	15.8	65.8	8082	1	US-08-306-691B-41	Sequence 41, Appl
c	28	15.8	65.8	8082	1	US-08-187-785-1	Sequence 1, Appli
c	29	15.8	65.8	8082	5	PCT-US93-06251-28	Sequence 28, Appl
	30	15.8	65.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	31	15.8	65.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	32	15.6	65.0	342	4	US-09-060-756-286	Sequence 286, App
c	33	15.6	65.0	50341	1	US-08-247-901C-1	Sequence 1, Appli
c	34	15.6	65.0	50341	2	US-09-075-904-1	Sequence 1, Appli
c	35	15.6	65.0	52297	4	US-09-426-436-1	Sequence 1, Appli
c	36	15.6	65.0	52297	4	US-08-705-557-1	Sequence 1, Appli
c	37	15.4	64.2	1752	3	US-08-941-445A-12	Sequence 12, Appl
c	38	15.4	64.2	2383	4	US-09-192-909-1	Sequence 1, Appli
c	39	15.4	64.2	2990	1	US-08-572-951-1	Sequence 1, Appli
c	40	15.4	64.2	3186	1	US-08-761-258-8	Sequence 8, Appli
c	41	15.4	64.2	3186	2	US-08-977-306-8	Sequence 8, Appli
	42	15.2	63.3	35	3	US-09-159-274-34	Sequence 34, Appl
	43	15.2	63.3	35	3	US-09-159-274-36	Sequence 36, Appl
	44	15.2	63.3	35	3	US-09-159-274-38	Sequence 38, Appl
	45	15.2	63.3	57	4	US-09-461-697-29	Sequence 29, Appl

ALIGNMENTS

```
RESULT 1
US-08-303-861-17/c
; Sequence 17, Application US/08303861
; Patent No. 6086902
;  GENERAL INFORMATION:
;    APPLICANT:  ZAMB, TIMOTHY
;    APPLICANT:  LIANG, XIAOPING
;    APPLICANT:  BABIUK, LORNE A.
;    TITLE OF INVENTION:  RECOMBINANT BOVINE HERPESVIRUS TYPE I
;    TITLE OF INVENTION:  VACCINES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/303,861
;      FILING DATE:  09-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PARK, FREDDIE K.
;      REGISTRATION NUMBER:  35,636
;      REFERENCE/DOCKET NUMBER:  29310-20020.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1499 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  390..1163
US-08-303-861-17
```

Query Match 72.5%; Score 17.4; DB 3; Length 1499;
Best Local Similarity 94.7%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       6 cgggcgcgaggaagcccgg 24
           |||||||||||| ||||||
Db     695 CGGGCGCGAGGACGCCCGG 677


RESULT   2
US-09-213-343-1/c
; Sequence 1, Application US/09213343
; Patent No. 6316252
; GENERAL INFORMATION:
;  APPLICANT: Harms, Jerome S.
;  APPLICANT: Splitter, Gary A.
;  TITLE OF INVENTION: Biotherapeutic Delivery System
;  FILE REFERENCE: 960296.95564
;  CURRENT APPLICATION NUMBER: US/09/213,343
;  CURRENT FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1499
;   TYPE: DNA
;   ORGANISM: Bovine herpesvirus 1
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (390)..(1166)
US-09-213-343-1


  Query Match             72.5%;  Score 17.4;  DB 4;  Length 1499;
  Best Local Similarity   94.7%;  Pred. No. 35;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 cgggcgcgaggaagcccgg 24
           |||||||||||| ||||||
Db     695 CGGGCGCGAGGACGCCCGG 677


RESULT   3
US-08-494-907-19
; Sequence 19, Application US/08494907
; Patent No. 5955298
;  GENERAL INFORMATION:
;    APPLICANT:  Thomashow, Linda S
;    APPLICANT:  Bangera, Mahalaxmi
;    APPLICANT:  Weller, David M
;    APPLICANT:  Cook, R. James
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Margaret A. Connor, USDA-ARS
;      STREET:  800 Buchanan Street
;      CITY:  Albany
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94710
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/494,907
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connor, Margaret A
;      REGISTRATION NUMBER:  30043
;      REFERENCE/DOCKET NUMBER:  0009.95
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 559-6067
;      TELEFAX:  (510) 559-5777
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  390 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..390
;      OTHER INFORMATION:  /note= "phlF, truncated, DNA
;      OTHER INFORMATION:  sequence.  SEQ ID NO:20 is translation (protein)
;      OTHER INFORMATION:  of SEQ ID NO:19.  "
US-08-494-907-19


  Query Match             67.5%;  Score 16.2;  DB 2;  Length 390;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ccgttcgggcgcgaggaagcc 21
           |||| | |||||||| |||||
Db     135 CCGTGCCGGCGCGAGCAAGCC 155


RESULT   4
PCT-US96-10986-19
; Sequence 19, Application PC/TUS9610986
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephan A. Pendorf, DOMINIK & STEIN
;      STREET:  600 N. West Shore Boulevard, Suite 1000
;      CITY:  Tampa
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33609
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10986
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pendorf, Stephan A.
;      REGISTRATION NUMBER:  32665
;      REFERENCE/DOCKET NUMBER:  A700.320
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (813) 289-2966
;      TELEFAX:  (813)289-2967
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  390 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..390
```

```
;       OTHER INFORMATION:  /note= "phlF, truncated, DNA
;       OTHER INFORMATION:  sequence.  SEQ ID NO:20 is translation (protein)
;       OTHER INFORMATION:  of SEQ ID NO:19.  "
PCT-US96-10986-19


  Query Match             67.5%;  Score 16.2;  DB 5;  Length 390;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcc 21
          |||| | |||||||| |||||
Db    135 CCGTGCCGGCGCGAGCAAGCC 155


RESULT   5
US-08-494-907-15
; Sequence 15, Application US/08494907
; Patent No. 5955298
;  GENERAL INFORMATION:
;    APPLICANT:  Thomashow, Linda S
;    APPLICANT:  Bangera, Mahalaxmi
;    APPLICANT:  Weller, David M
;    APPLICANT:  Cook, R. James
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Margaret A. Connor, USDA-ARS
;      STREET:  800 Buchanan Street
;      CITY:  Albany
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94710
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/494,907
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connor, Margaret A
;      REGISTRATION NUMBER:  30043
;      REFERENCE/DOCKET NUMBER:  0009.95
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 559-6067
;      TELEFAX:  (510) 559-5777
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  606 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..606
;      OTHER INFORMATION:  /note= "phlF DNA sequence.  SEQ ID
;      OTHER INFORMATION:  NO:16 is translation (protein) of SEQ ID NO:15."
US-08-494-907-15


  Query Match             67.5%;  Score 16.2;  DB 2;  Length 606;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcc 21
          |||| | |||||||| |||||
Db    135 CCGTGCCGGCGCGAGCAAGCC 155


RESULT   6
PCT-US96-10986-15
; Sequence 15, Application PC/TUS9610986
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephan A. Pendorf, DOMINIK & STEIN
;      STREET:  600 N. West Shore Boulevard, Suite 1000
;      CITY:  Tampa
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33609
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10986
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pendorf, Stephan A.
;      REGISTRATION NUMBER:  32665
;      REFERENCE/DOCKET NUMBER:  A700.320
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (813) 289-2966
;      TELEFAX:  (813)289-2967
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  606 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..606
;      OTHER INFORMATION:  /note= "phlF DNA sequence.  SEQ ID
;      OTHER INFORMATION:  NO:16 is translation (protein) of SEQ ID NO:15."
PCT-US96-10986-15


  Query Match             67.5%;  Score 16.2;  DB 5;  Length 606;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcc 21
          |||| | |||||||| |||||
Db    135 CCGTGCCGGCGCGAGCAAGCC 155


RESULT   7
US-08-494-907-4
; Sequence 4, Application US/08494907
; Patent No. 5955298
;  GENERAL INFORMATION:
;    APPLICANT:  Thomashow, Linda S
;    APPLICANT:  Bangera, Mahalaxmi
;    APPLICANT:  Weller, David M
;    APPLICANT:  Cook, R. James
;    TITLE OF INVENTION:  Sequences for Production of
```

```
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Margaret A. Connor, USDA-ARS
;      STREET:  800 Buchanan Street
;      CITY:  Albany
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94710
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/494,907
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connor, Margaret A
;      REGISTRATION NUMBER:  30043
;      REFERENCE/DOCKET NUMBER:  0009.95
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 559-6067
;      TELEFAX:  (510) 559-5777
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6170 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (4285..5076)
;      OTHER INFORMATION:  /note= "phlA, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (2606..3313)
;      OTHER INFORMATION:  /note= "phlB, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (3085..4251)
;      OTHER INFORMATION:  /note= "phlC, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (1398..2444)
;      OTHER INFORMATION:  /note= "phlD, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3514..4767
;      OTHER INFORMATION:  /note= "phlR, transcribed from
;      OTHER INFORMATION:  left to right"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (2..1270)
;      OTHER INFORMATION:  /note= "phlE, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5781..6170
;      OTHER INFORMATION:  /note= "phlF, truncated,
;      OTHER INFORMATION:  transcribed from left to right"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..6170
;      OTHER INFORMATION:  /note= "SEQ ID NO:4 contains genes
;      OTHER INFORMATION:  involved in synthesis, and modulation of synthesis
;      OTHER INFORMATION:  of Phl."
US-08-494-907-4


  Query Match             67.5%;  Score 16.2;  DB 2;  Length 6170;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ccgttcgggcgcgaggaagcc 21
           |||| | ||||||||| |||||
Db    5915 CCGTGCCGGCGCGAGCAAGCC 5935


RESULT   8
PCT-US96-10986-4
; Sequence 4, Application PC/TUS9610986
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephan A. Pendorf, DOMINIK & STEIN
;      STREET:  600 N. West Shore Boulevard, Suite 1000
;      CITY:  Tampa
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33609
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/10986
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pendorf, Stephan A.
;      REGISTRATION NUMBER:  32665
;      REFERENCE/DOCKET NUMBER:  A700.320
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (813) 289-2966
;      TELEFAX:  (813)289-2967
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6170 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (4285..5076)
;      OTHER INFORMATION:  /note= "phlA, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (2606..3313)
;      OTHER INFORMATION:  /note= "phlB, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (3085..4251)
;      OTHER INFORMATION:  /note= "phlC, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
```

```
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (1398..2444)
;       OTHER INFORMATION:  /note= "phlD, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3514..4767
;       OTHER INFORMATION:  /note= "phlR, transcribed from
;       OTHER INFORMATION:  left to right"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (2..1270)
;       OTHER INFORMATION:  /note= "phlE, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5781..6170
;       OTHER INFORMATION:  /note= "phlF, truncated,
;       OTHER INFORMATION:  transcribed from left to right"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..6170
;       OTHER INFORMATION:  /note= "SEQ ID NO:4 contains genes
;       OTHER INFORMATION:  involved in synthesis, and modulation of synthesis
;       OTHER INFORMATION:  of Phl."
PCT-US96-10986-4


  Query Match             67.5%;  Score 16.2;  DB 5;  Length 6170;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcc 21
          |||| | |||||||| |||||
Db   5915 CCGTGCCGGCGCGAGCAAGCC 5935


RESULT   9
US-08-494-907-3
; Sequence 3, Application US/08494907
; Patent No. 5955298
;  GENERAL INFORMATION:
;    APPLICANT:  Thomashow, Linda S
;    APPLICANT:  Bangera, Mahalaxmi
;    APPLICANT:  Weller, David M
;    APPLICANT:  Cook, R. James
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Margaret A. Connor, USDA-ARS
;      STREET:  800 Buchanan Street
;      CITY:  Albany
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94710
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/494,907
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connor, Margaret A
;      REGISTRATION NUMBER:  30043
;      REFERENCE/DOCKET NUMBER:  0009.95
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 559-6067
;      TELEFAX:  (510) 559-5777
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6387 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (4285..5076)
;      OTHER INFORMATION:  /note= "phlA, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (2606..3313)
;      OTHER INFORMATION:  /note= "phlB, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (3085..4251)
;      OTHER INFORMATION:  /note= "phlC, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (1398..2444)
;      OTHER INFORMATION:  /note= "phlD, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3514..4767
;      OTHER INFORMATION:  /note= "phlR, transcribed from left
;      OTHER INFORMATION:  to right"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (2..1270)
;      OTHER INFORMATION:  /note= "phlE, transcribed from
;      OTHER INFORMATION:  right to left"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5781..6386
;      OTHER INFORMATION:  /note= "phlF, transcribed from left
;      OTHER INFORMATION:  to right"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..6387
;      OTHER INFORMATION:  /note= "SEQ ID NO:3 contains genes
;      OTHER INFORMATION:  involved in synthesis and modulation of synthesis
;      OTHER INFORMATION:  of Phl."
US-08-494-907-3


  Query Match             67.5%;  Score 16.2;  DB 2;  Length 6387;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ccgttcgggcgcgaggaagcc 21
          |||| | |||||||| |||||
Db   5915 CCGTGCCGGCGCGAGCAAGCC 5935


RESULT  10
PCT-US96-10986-3
; Sequence 3, Application PC/TUS9610986
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  Sequences for Production of
;    TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Stephan A. Pendorf, DOMINIK & STEIN
;      STREET:  600 N. West Shore Boulevard, Suite 1000
```

```
;       CITY:  Tampa
;       STATE:  FL
;       COUNTRY:  USA
;       ZIP:  33609
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/10986
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pendorf, Stephan A.
;       REGISTRATION NUMBER:  32665
;       REFERENCE/DOCKET NUMBER:  A700.320
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (813) 289-2966
;       TELEFAX:  (813)289-2967
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6387 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Pseudomonas fluorescens
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (4285..5076)
;       OTHER INFORMATION:  /note= "phlA, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (2606..3313)
;       OTHER INFORMATION:  /note= "phlB, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (3085..4251)
;       OTHER INFORMATION:  /note= "phlC, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (1398..2444)
;       OTHER INFORMATION:  /note= "phlD, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3514..4767
;       OTHER INFORMATION:  /note= "phlR, transcribed from left
;       OTHER INFORMATION:  to right"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (2..1270)
;       OTHER INFORMATION:  /note= "phlE, transcribed from
;       OTHER INFORMATION:  right to left"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5781..6386
;       OTHER INFORMATION:  /note= "phlF, transcribed from left
;       OTHER INFORMATION:  to right"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..6387
;       OTHER INFORMATION:  /note= "SEQ ID NO:3 contains genes
;       OTHER INFORMATION:  involved in synthesis and modulation of synthesis
;       OTHER INFORMATION:  of Phl."
PCT-US96-10986-3
```

```
  Query Match           67.5%;  Score 16.2;  DB 5;  Length 6387;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ccgttcgggcgcgaggaagcc 21
           |||| | ||||||||| |||||
Db    5915 CCGTGCCGGCGCGAGCAAGCC 5935


RESULT  11
US-08-188-281B-8
; Sequence 8, Application US/08188281B
; Patent No. 5610009
;  GENERAL INFORMATION:
;     APPLICANT:  WATANABE, SHINICHI
;     APPLICANT:  YAMAGUCHI, JULIE
;     APPLICANT:  DESAI, SURESH M.
;     APPLICANT:  DEVARE, SUSHIL G.
;     TITLE OF INVENTION:  MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;     TITLE OF INVENTION:  ENVELOPE GENES
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;       STREET:  ONE ABBOTT PARK ROAD
;       CITY:  ABBOTT PARK
;       STATE:  IL
;       COUNTRY:  USA
;       ZIP:  60064-3500
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/188,281B
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  POREMBSKI, PRISCILLA E.
;       REGISTRATION NUMBER:  33,207
;       REFERENCE/DOCKET NUMBER:  5521.US.01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  708-937-6365
;       TELEFAX:  708-938-2623
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7106 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  922..2022
US-08-188-281B-8


  Query Match           67.5%;  Score 16.2;  DB 1;  Length 7106;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 ttcgggcgcgaggaagcccgg 24
           |||||||||||| |  ||||||
Db     912 TTCGGGCGCGATGCTGCCCGG 932


RESULT  12
US-08-453-552-5
; Sequence 5, Application US/08453552
; Patent No. 5667992
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  CASEY, JAMES M.
;    APPLICANT:  BODE, SUZANNE L.
;    APPLICANT:  ZECK, BILLY J.
;    APPLICANT:  YAMAGUCHI, JULIE
;    APPLICANT:  FRAIL, DONALD E.
;    APPLICANT:  DESAI, SURESH M.
;    APPLICANT:  DEVARE, SUSHIL G.
;    TITLE OF INVENTION:  MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,552
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5131.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7106 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  922..2022
US-08-453-552-5


  Query Match             67.5%;  Score 16.2;  DB 1;  Length 7106;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 ttcgggcgcgaggaagcccgg 24
           |||||||||||  |  ||||||
Db     912 TTCGGGCGCGATGCTGCCCGG 932


RESULT  13
US-08-710-637-5
; Sequence 5, Application US/08710637
; Patent No. 5854001
;  GENERAL INFORMATION:
;    APPLICANT:  CASEY, JAMES M.
;    APPLICANT:  BODE, SUZANNE L.
;    APPLICANT:  ZECK, BILLY J.
;    APPLICANT:  YAMAGUCHI, JULIE
;    APPLICANT:  FRAIL, DONALD E.
;    APPLICANT:  DESAI, SURESH M.
;    APPLICANT:  DEVARE, SUSHIL G.
;    TITLE OF INVENTION:  MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/710,637
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/144,099
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/830,024
;      FILING DATE:  01-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5131.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7106 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  922..2022
US-08-710-637-5


  Query Match             67.5%;  Score 16.2;  DB 2;  Length 7106;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 ttcgggcgcgaggaagcccgg 24
           |||||||||||  |  ||||||
Db     912 TTCGGGCGCGATGCTGCCCGG 932


RESULT  14
PCT-US93-00907-5
; Sequence 5, Application PC/TUS9300907
;  GENERAL INFORMATION:
;    APPLICANT:  CASEY, JAMES M.
;    APPLICANT:  BODE, SUZANNE L.
;    APPLICANT:  ZECK, BILLY J.
;    APPLICANT:  YAMAGUCHI, JULIE
;    APPLICANT:  FRAIL, DONALD E.
;    APPLICANT:  DESAI, SURESH M.
;    APPLICANT:  DEVARE, SUSHIL G.
;    TITLE OF INVENTION:  MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/00907
;      FILING DATE:  19930129
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5131.PC.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7106 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  922..2022
PCT-US93-00907-5


  Query Match             67.5%;  Score 16.2;  DB 5;  Length 7106;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 ttcgggcgcgaggaagcccgg 24
           ||||||||||| |  ||||||
Db     912 TTCGGGCGCGATGCTGCCCGG 932


RESULT  15
PCT-US94-07280-8
; Sequence 8, Application PC/TUS9407280
;  GENERAL INFORMATION:
;    APPLICANT:  WATANABE, SHINICHI
;    APPLICANT:  YAMAGUCHI, JULIE
;    APPLICANT:  DESAI, SURESH M.
;    APPLICANT:  DEVARE, SUSHIL G.
;    TITLE OF INVENTION:  MAMMALIAN EXPRESSION SYSTEMS FOR HCV
;    TITLE OF INVENTION:  ENVELOPE GENES
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07280
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5521.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7106 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  922..2022
PCT-US94-07280-8


  Query Match             67.5%;  Score 16.2;  DB 5;  Length 7106;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 ttcgggcgcgaggaagcccgg 24
           ||||||||||| |  ||||||
Db     912 TTCGGGCGCGATGCTGCCCGG 932


Search completed: June 23, 2002, 15:09:16
Job time: 78580 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 06:32:15 ; Search time 2161.72 Seconds
(without alignments)
212.971 Million cell updates/sec

Title: US-09-747-514A-4
Perfect score: 22
Sequence: 1 gaagcggcgtaggccgggggtc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	22	100.0	3123	1	PAU38241	U38241 Pseudomonas
c 2	22	100.0	13075	1	AE004946	AE004946 Pseudomon
c 3	17.8	80.9	91686	2	AC094113	AC094113 Rattus no
c 4	17.4	79.1	10968	1	AE002062	AE002062 Deinococc
5	17.4	79.1	44273	3	AY028171	AY028171 Leishmani
6	17.2	78.2	2317	1	HSU53365	U53365 Halobacteri
c 7	17.2	78.2	4287	1	HSHTPIVAV	X95589 H.salinariu
c 8	17.2	78.2	8807	3	DVHB	X15359 Drosophila
9	17.2	78.2	10784	1	AE005080	AE005080 Halobacte
c 10	17.2	78.2	28969	2	AC094368	AC094368 Rattus no
11	17.2	78.2	38793	1	MSGB27CS	L78817 Mycobacteri
c 12	17.2	78.2	64189	2	AC015511	AC015511 Homo sapi
c 13	17.2	78.2	65467	2	AC017402	AC017402 Drosophil
14	17.2	78.2	100814	2	AP003929	AP003929 Oryza sat
15	17.2	78.2	104278	3	AC005269	AC005269 Drosophil
c 16	17.2	78.2	118995	9	AC005368	AC005368 Homo sapi
c 17	17.2	78.2	121011	2	AC103003	AC103003 Rattus no
18	17.2	78.2	121788	2	AC022420	AC022420 Homo sapi
19	17.2	78.2	126682	9	AC008649	AC008649 Homo sapi
c 20	17.2	78.2	135259	9	AC004590	AC004590 Homo sapi
c 21	17.2	78.2	175504	2	AC073149	AC073149 Homo sapi
22	17.2	78.2	179927	2	AC021491	AC021491 Homo sapi
c 23	17.2	78.2	182726	3	AC008002	AC008002 Drosophil
24	17.2	78.2	191311	2	AC091486	AC091486 Homo sapi
c 25	17.2	78.2	197744	2	AC008439	AC008439 Homo sapi
26	17.2	78.2	205516	9	AC008781	AC008781 Homo sapi
27	17.2	78.2	273785	1	SME591793	AL591793 Sinorhizo
28	17.2	78.2	302473	3	AE003589	AE003589 Drosophil
29	17.2	78.2	348450	1	MLEPRTN4	AL583920 Mycobacte
30	16.8	76.4	96133	2	AC105465	AC105465 Rattus no
c 31	16.8	76.4	105416	2	AC099263	AC099263 Rattus no
c 32	16.8	76.4	123062	2	AC099359	AC099359 Rattus no
c 33	16.8	76.4	159587	2	AC098547	AC098547 Rattus no
c 34	16.8	76.4	161798	2	AP003528	AP003528 Oryza sat
35	16.8	76.4	165202	2	AP003491	AP003491 Oryza sat
c 36	16.8	76.4	173221	2	AC106194	AC106194 Rattus no
37	16.8	76.4	196521	2	AC007873	AC007873 Homo sapi
38	16.4	74.5	10610	1	AE002029	AE002029 Deinococc
39	16.4	74.5	30000	6	AX250261	AX250261 Sequence
c 40	16.4	74.5	77670	7	AF222060	AF222060 Halovirus
41	16.4	74.5	164839	2	AC105823	AC105823 Rattus no
c 42	16.2	73.6	635	9	HSA324213	AJ324213 Homo sapi
43	16.2	73.6	820	33	AC076948	Ac076948 Giardia i
c 44	16.2	73.6	1454	9	HSM801746	AL136778 Homo sapi
45	16.2	73.6	1768	9	AK027396	AK027396 Homo sapi

ALIGNMENTS

RESULT 1
PAU38241/c
LOCUS PAU38241 3123 bp DNA linear BCT 05-OCT-1996
DEFINITION Pseudomonas aeruginosa orotate phophoribosyl transferase (pyrE), catabolite repression control protein (crc) and RNasePH (rph) genes, complete cds.
ACCESSION U38241 L12038
VERSION U38241.1 GI:1079660
KEYWORDS .
SOURCE Pseudomonas aeruginosa strain=PAO1.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 3123)
AUTHORS MacGregor,C.H., Arora,S.K., Hager,P.W., Dail,M.B. and Phibbs,P.V. Jr.
TITLE The nucleotide sequence of the Pseudomonas aeruginosa pyrE-crc-rph region and the purification of the crc gene product
JOURNAL J. Bacteriol. 178 (19), 5627-5635 (1996)
MEDLINE 96421989
REFERENCE 2 (bases 1 to 3123)
AUTHORS Hager,P.W. and Phibbs,P.V. Jr.
TITLE Direct Submission

```
  JOURNAL   Submitted (11-OCT-1995) Paul W. Hager, Microbiology & Immunology,
            East Carolina University, Greenville, NC 27858, USA
COMMENT     On Nov 29, 1995 this sequence version replaced gi:496210.
FEATURES             Location/Qualifiers
     source          1. .3123
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
                     /map="11 min"
     gene            complement(414. .1055)
                     /gene="pyrE"
     CDS             complement(414. .1055)
                     /gene="pyrE"
                     /EC_number="2.4.2.10"
                     /function="catalytic activity: orotidine-5'-phosphate +
                     pyrophosphate = orotate + 5-phospho-alpha-d-ribose
                     1-diphosphate; Method: conceptual translation supplied by
                     author."
                     /codon_start=1
                     /transl_table=11
                     /product="orotate phosphoribosyl transferase"
                     /protein_id="AAC44427.1"
                     /db_xref="GI:1079661"
                     /translation="MQAYQRDFIRFAIERGVLRFGEFTLKSGRTSPYFFNAGLFDSGL
                     ALARLGRFYAEAVIDSGIDFDVLFGPAYKGIPLAATTAVALAEQHQRDLPWCFNRKEA
                     KEHGEGGTLVGAPLSGRVLIIDDVITAGTAIREVMQIIDAQGARAAGVLIALNRQERG
                     KGELSAIQEVERDFGMPVVSIVSLEQVLEYLAEDAELKKHLPAVEAYRAQYGI"
     gene            1136. .1915
                     /gene="crc"
     CDS             1136. .1915
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAC44428.1"
                     /db_xref="GI:1079662"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(2394. .3113)
                     /gene="rph"
     CDS             complement(2394. .3113)
                     /gene="rph"
                     /note="similar to ribonuclease PH"
                     /codon_start=1
                     /transl_table=11
                     /product="RNase PH"
                     /protein_id="AAC44429.1"
                     /db_xref="GI:1079663"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
                     ESGVPRFLKGQGQGWLTAEYGMLRRSTGERNQREASRGKQGGRTLEIQRLIDRSLRAA
                     LDLSKLGENTLYIDCDVIQTDGGTRTASITGATVSLIDALAVLKKRVALKGNPLKQIV
                     AAVSVGIYQGVPVLDLDYLEDSAAESDLNVVMTDCRRFIEVQGTAEGAPFRPAELNAM
                     LELAQQGMQELVRTQRAALAE"
BASE COUNT      551 a   1020 c    989 g    563 t
ORIGIN

  Query Match           100.0%;  Score 22;  DB 1;  Length 3123;
  Best Local Similarity  100.0%;  Pred. No. 34;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           ||||||||||||||||||||||
Db    1828 GAAGCGGCGTAGGCCGGGGGTC 1807

RESULT   2
AE004946/c
LOCUS       AE004946               13075 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 507 of 529 of the complete
            genome.
ACCESSION   AE004946 AE004091
VERSION     AE004946.1  GI:9951650
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 13075)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .13075
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            71. .850
                     /gene="crc"
                     /note="PA5332"
     CDS             71. .850
                     /gene="crc"
                     /codon_start=1
                     /transl_table=11
                     /product="catabolite repression control protein"
                     /protein_id="AAG08717.1"
                     /db_xref="GI:9951651"
                     /translation="MRIISVNVNGIQAAAERGLLSWLQAQNADVICLQDTRASAFDLD
                     DPSFQLDGYFLYACDAELPEQGGVALYSRLQPKAVISGLGFETADRYGRYLQADFDKV
                     SIATLLLPSGQSGDESLNQKFKFMDDFTHYLSKQRRKRREYIYCGSLYVAHQKMDVKN
                     WRECQQMPGFLAPERAWLDEVFGNLGYADALREVSREGDQFSWWPDSEQAEMLNLGWR
                     FDYQVLTPGLRRFVRNAKLPRQPRFSQHAPLIVDYDWQLSI"
     gene            complement(909. .1280)
                     /gene="PA5333"
     CDS             complement(909. .1280)
                     /gene="PA5333"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG08718.1"
                     /db_xref="GI:9951652"
                     /translation="MNDVSEEVAATPGQEARQWAMFCHFAAFIGLVFPFGSLLGPLIV
                     WQLKRESDPFVDAQGKEALNFQITVALAMFVSFLLMLVVIGFFLLGLVCLAALVLTII
                     AGIKANEGKAYRYPFCLRLIK"
     gene            complement(1327. .2046)
                     /gene="rph"
                     /note="PA5334"
     CDS             complement(1327. .2046)
                     /gene="rph"
                     /codon_start=1
                     /transl_table=11
                     /product="ribonuclease PH"
                     /protein_id="AAG08719.1"
                     /db_xref="GI:9951653"
                     /translation="MNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTVSA
```

Comp

ESGVPRFLKGQGQGWLTAEYGMLPRSTGERNQREASRGKQGGRTLEIQRLIGRSLRAA
LDLSKLGENTLYIDCDVIQADGGTRTASITGATVALIDALAVLKKRGALKGNPLKQMV
AAVSVGIYQGVPVLDLDYLEDSAAETDLNVVMTDAGGFIEVQGTAEGAPFRPAELNAM
LELAQQGMQELFELQRAALAE"

gene 2227. .3090
/gene="PA5335"

CDS 2227. .3090
/gene="PA5335"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08720.1"
/db_xref="GI:9951654"
/translation="MVHSMTAFARVEQAGTHGTLSWELRSVNHRYLEPHLRLPDAFRD
LEGAVREALRQGLSRGKVECTLRFAEETAGKSLQVDQERARQLVAAAEGVAALIRQPA
PLDPLAVLAWPGVLVADSADPQALNAAALEAFGQALEQLKAGRSREGLELAKLLNDRL
DAMLVEVGNLRELVPTMLANQRQKILDRFAELKAELDPQRLEQELVLLAQKSDVAEEL
DRLATHVGEVRRVLKAGGAAGRRLDFLMQELNREANTLGSKAFDPRSTQAAVNLKVLI
EQMREQVQNIE"

gene 3149. .3760
/gene="gmk"
/note="PA5336"

CDS 3149. .3760
/gene="gmk"
/codon_start=1
/transl_table=11
/product="guanylate kinase"
/protein_id="AAG08721.1"
/db_xref="GI:9951655"
/translation="MSGTLYIVSAPSGAGKTSLVKALLDAAPEVRVSVSHTTRGMRPG
EVDGVNYHFTSREEFLAMLERNEFLEHAEVFGNLYGTSQRWVEKTLAEGLDLILEIDW
QGAQQVRRLMPEAQSIFILPPSQEALRQRLTNRGQDSDEVIERRMREAVSEMSHYVEY
DHLVINDDFAHALDDLKAIFRARQLRQDAQQQRHAELLGRLLA"

gene 3849. .4115
/gene="rpoZ"
/note="PA5337"

CDS 3849. .4115
/gene="rpoZ"
/codon_start=1
/transl_table=11
/product="RNA polymerase omega subunit"
/protein_id="AAG08722.1"
/db_xref="GI:9951656"
/translation="MARVTVEDCLDNVDNRFELVMLATKRARQLATGGKEPKVAWEND
KPTVVALREIASGLVDENVVQQEDIVEDEPLFAAFDDEANTEAL"

gene 4181. .6286
/gene="spoT"
/note="PA5338"

CDS 4181. .6286
/gene="spoT"
/codon_start=1
/transl_table=11
/product="guanosine-3',5'-bis(diphosphate)
3'-pyrophosphohydrolase"
/protein_id="AAG08723.1"
/db_xref="GI:9951657"
/translation="MPGIDAFADRLSSYLGPDQVNLVRRAYYYAEQAHDGQRRRSGEP
YVTHPLAVSNILADMHMDHQSLMAAMLHDVIEDTGIAKEALTAQFGETVSELVDGVSK
LTQMNFETKAEAQAENFQKMAMAMARDIRVILVKLADRLHNMRTLEVLSGEKRRRIAK
ETLEIYAPIANRLGMHTMRVEFEDLGFKAMHPMRAERIRQAVKKARGNRREIVGKIQE
SLVNCLAREGMEGQVIGREKHLYSIYQKMRGKRKAFNEIMDVYAFRIIVDKVDTCYRV
LGAVHSLYKPFPGRFKDYIAIPKANGYQSLHTTLFGMHGVPIEIQIRTREMEEMANHG
IAAHWLYKSNDETPKGTHARARQWVKGVLELQQRAGNSLEFIENVKIDLFPDEVYVFT
PKGRIMELPKGSTAVDFAYAVHTDVGNSCIACRINRRLAPLSEPLQSGQTVEIVTAPG
ARPNPAWLSFVVTGKARTHIRHALKLQRRSESINLGERLLNKTLTGFDSHLEKIPQER
IQAVLAEYRMEVFEDLLEEIGLGNRMAYVVARRLLSSEGEEAPSAEGPLAIRGTEGLV
LSYAKCCTPIPGDPIVGHLSAGKGMVVHLESCKNISEVRHNPEKCIQLSWAKDVTGEF
NVELRVELEHQRGLIALLATSVNAADGNIEKISMDERDGRISVVQLVVSVHDRVHLAR
VIKKLRALKGVIRITRLRS"

gene 6348. .6728
/gene="PA5339"

CDS 6348. .6728
/gene="PA5339"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08724.1"
/db_xref="GI:9951658"
/translation="MTKTVIHTDNAPAAIGTYSQAIKAGNTVYVSGQIPLDPKTMELV
EGFEEQTVQVFENLKAVIEASGGYLHDVAKLNIFLTDLSHFAKVNEIMGRYFTQPYPA
RAAIGVAALPKGAQVEMDAIVVLE"

gene 6783. .7514
/gene="PA5340"

CDS 6783. .7514
/gene="PA5340"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08725.1"
/db_xref="GI:9951659"
/translation="MRRVIFLAAAATLLAGCAGTADPSGTWINQAAIDAASKDGKLRE
ALLAYGPNLEWKLDSKAGEATFSNGFELGEGTLSKSDDEHWKVAFYGDDNQESLELDG
KELIQQASANGPEQRFRRLDPQPAASSPAGSGFERALYGSYLKGSWKIREGQGQGGKV
EFQANGLVSGLPGAERYALCLAGDCAAMSGDNDSIWLQQGNRGRELLFSLDDDELQLF
EAVNTAGANEMPSYVPGKRVWLLER"

gene complement(7521. .8141)
/gene="PA5341"

CDS complement(7521. .8141)
/gene="PA5341"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08726.1"
/db_xref="GI:9951660"
/translation="MFAAFVLVASTHFAALLSPGPDFFLLVRAALLRGRRQADGVAAG
IALANLLSMLLVLGLLASVPDSAHSALRWLQALGGLYFLWLAGQALLAQRRLEMPAQR
DVPSRGGYLRGLRDGLLASSLNPKLPIFYAGLFGVLARFSLPGWALALCLAWMSLAVL
CWDMALVRLLDRPRWRGWLQRRVGALDRLCGVLLLALGGWLVLAAL"

gene complement(8208. .9008)
/gene="PA5342"

CDS complement(8208. .9008)
/gene="PA5342"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG08727.1"
/db_xref="GI:9951661"
/translation="MAANRLLARPWLPGVELFHADFSGQAFGRHSHSAFAIGSILHGV
GGYQCRGRRHALPAGTLSLMNPEEPHTGHAESERLVYRMLYVEETRLLGLLGRKQLPR
GFGELNPADDGSVAAGLAWLAEDFQRSDALGLESRLLALLEQVFVRHGGLRPAAAAPR
DGGTTARLRDYLEAHATEALGLDQLAQLLQRHPRHLIEAFRRAYGVPPHTYQLQCRVR
QAKRLLLEERPLAELAQHLGFYDQAHFSTTFRRFTGVTPGQFRRSAKT"

gene complement(9008. .9859)
/gene="PA5343"

CDS complement(9008. .9859)

Query Match 100.0%; Score 22; DB 1; Length 13075;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gaagcggcgtaggccgggggtc 22
||||||||||||||||||||||
Db 763 GAAGCGGCGTAGGCCGGGGGTC 742

COMP

RESULT 3
AC094113/c
LOCUS AC094113 91686 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-2K1, *** SEQUENCING IN PROGRESS ***, 51 unordered pieces.
ACCESSION AC094113
VERSION AC094113.2 GI:17940822
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 91686)

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 91686)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Dec 20, 2001 this sequence version replaced gi:15633017.

---------- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

---------- Project Information

Center project name: GAAB

Center clone name: CH230-2K1

---------- Summary Statistics

Assembly program: Phrap; version 0.990329First call to findPhrapList

Consensus quality: 60493 bases at least Q40

Consensus quality: 68816 bases at least Q30

Consensus quality: 75216 bases at least Q20

Estimated insert size: 52587; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation

---

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	6013: contig of 6013 bp in length
*	6014	6113: gap of unknown length
*	6114	11721: contig of 5608 bp in length
*	11722	11821: gap of unknown length
*	11822	14786: contig of 2965 bp in length
*	14787	14886: gap of unknown length
*	14887	17633: contig of 2747 bp in length
*	17634	17733: gap of unknown length
*	17734	20201: contig of 2468 bp in length
*	20202	20301: gap of unknown length
*	20302	23127: contig of 2826 bp in length
*	23128	23227: gap of unknown length
*	23228	24685: contig of 1458 bp in length
*	24686	24785: gap of unknown length
*	24786	26709: contig of 1924 bp in length
*	26710	26809: gap of unknown length
*	26810	28571: contig of 1762 bp in length
*	28572	28671: gap of unknown length
*	28672	29792: contig of 1121 bp in length
*	29793	29892: gap of unknown length
*	29893	31227: contig of 1335 bp in length
*	31228	31327: gap of unknown length
*	31328	33334: contig of 2007 bp in length
*	33335	33434: gap of unknown length
*	33435	35058: contig of 1624 bp in length
*	35059	35158: gap of unknown length
*	35159	36583: contig of 1425 bp in length
*	36584	36683: gap of unknown length
*	36684	38115: contig of 1432 bp in length
*	38116	38215: gap of unknown length
*	38216	39613: contig of 1398 bp in length
*	39614	39713: gap of unknown length
*	39714	41308: contig of 1595 bp in length
*	41309	41408: gap of unknown length
*	41409	42858: contig of 1450 bp in length
*	42859	42958: gap of unknown length
*	42959	44357: contig of 1399 bp in length
*	44358	44457: gap of unknown length
*	44458	46699: contig of 2242 bp in length
*	46700	46799: gap of unknown length
*	46800	48262: contig of 1463 bp in length
*	48263	48362: gap of unknown length
*	48363	50195: contig of 1833 bp in length
*	50196	50295: gap of unknown length
*	50296	51912: contig of 1617 bp in length
*	51913	52012: gap of unknown length
*	52013	53341: contig of 1329 bp in length
*	53342	53441: gap of unknown length
*	53442	54750: contig of 1309 bp in length
*	54751	54850: gap of unknown length
*	54851	57276: contig of 2426 bp in length
*	57277	57376: gap of unknown length
*	57377	58467: contig of 1091 bp in length
*	58468	58567: gap of unknown length
*	58568	59618: contig of 1051 bp in length
*	59619	59718: gap of unknown length
*	59719	60919: contig of 1201 bp in length
*	60920	61019: gap of unknown length
*	61020	62112: contig of 1093 bp in length
*	62113	62212: gap of unknown length
*	62213	63324: contig of 1112 bp in length
*	63325	63424: gap of unknown length
*	63425	64773: contig of 1349 bp in length
*	64774	64873: gap of unknown length
*	64874	66366: contig of 1493 bp in length
*	66367	66466: gap of unknown length
*	66467	67872: contig of 1406 bp in length

```
              *    67873    67972: gap of unknown length
              *    67973    69389: contig of 1417 bp in length
              *    69390    69489: gap of unknown length
              *    69490    71217: contig of 1728 bp in length
              *    71218    71317: gap of unknown length
              *    71318    72371: contig of 1054 bp in length
              *    72372    72471: gap of unknown length
              *    72472    73631: contig of 1160 bp in length
              *    73632    73731: gap of unknown length
              *    73732    74871: contig of 1140 bp in length
              *    74872    74971: gap of unknown length
              *    74972    76348: contig of 1377 bp in length
              *    76349    76448: gap of unknown length
              *    76449    77823: contig of 1375 bp in length
              *    77824    77923: gap of unknown length
              *    77924    79539: contig of 1616 bp in length
              *    79540    79639: gap of unknown length
              *    79640    81122: contig of 1483 bp in length
              *    81123    81222: gap of unknown length
              *    81223    82241: contig of 1019 bp in length
              *    82242    82341: gap of unknown length
              *    82342    83562: contig of 1221 bp in length
              *    83563    83662: gap of unknown length
              *    83663    84681: contig of 1019 bp in length
              *    84682    84781: gap of unknown length
              *    84782    86335: contig of 1554 bp in length
              *    86336    86435: gap of unknown length
              *    86436    87812: contig of 1377 bp in length
              *    87813    87912: gap of unknown length
              *    87913    89049: contig of 1137 bp in length
              *    89050    89149: gap of unknown length
              *    89150    90468: contig of 1319 bp in length
              *    90469    90568: gap of unknown length
              *    90569    91686: contig of 1118 bp in length.
FEATURES             Location/Qualifiers
     source          1. .91686
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-2K1"
BASE COUNT    25253 a  17757 c  17593 g  25947 t   5136 others
ORIGIN


  Query Match           80.9%;  Score 17.8;  DB 2;  Length 91686;
  Best Local Similarity   90.5%;  Pred. No. 7.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggt 21
           ||||||||||| | |||||||
Db   57092 GAAGCGGCGTAAGACGGGGGT 57072


RESULT   4
AE002062/c
LOCUS       AE002062               10968 bp    DNA     linear   BCT 22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 199 of 229 of the complete
            chromosome 1.
ACCESSION   AE002062 AE000513
VERSION     AE002062.1  GI:6460110    .
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcaceae;
            Deinococcus.
REFERENCE   1  (bases 1 to 10968)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 10968)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10968
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            85. .2757
                     /gene="DR2300"
     CDS             85. .2757
                     /gene="DR2300"
                     /note="similar to SP:P74941 PID:1565288 percent identity:
                     68.33; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="alanyl-tRNA synthetase"
                     /protein_id="AAF11848.1"
                     /db_xref="GI:6460111"
                     /translation="MTAPLTTAEIRERYLYFFETKGHLRLPSHSTIAPDPTTLFTVAG
                     MQPFKENFMGAPAVFDGQPSKRVTTAQKCVRVGDIENVGRTRRHLSLFEMMGNFSFGD
                     YFKRDAIHWAWEFLTGPEWMGMDKDKMYVTVYKDDDEAFGYWTQDIGLPAEHIHRFDA
                     DENFWPANAPLEGPNGPCGPCSEIYYDRGENYGDDTWGDYYQTRESARFLEVWNLVFP
                     QYDRQDLDASGQPVLKDLPFKNIDTGMGLERVASVVQDVPDFYSNDVFKPIVERVAEL
                     SGKPYEGETSVSHRVVAEHIRSVSMIVADGTAFSNTGRGYTARKIMRRAIRHGYMLGL
                     REPQLYRLVELVVDSMGGAYPELQHNQSRVEASVRAEEEQFLKTLEGGIKRLGGLLSG
                     MEKGSTLVGKDAFELYDTYGFPVDLTKEIAEEYGVSVDEAGYAESLEHAQEIARAGSK
                     YGKSELFGGHQEALDGLPATQFVGYDQTSGDGQVLALLSGGERLSHLPAGSEATVVLS
                     QTPFYGEGGGEVGDTGRLEWDGGAGIVRDTQKTKQGVFLHDVLVEQGELKEGTRVRGV
                     VSPERAAIQRHHTATHLLHAALRAVLGGGVQQKGSRVAADQLRFDFSHGAAMSAEEIA
                     QVETLVSRWVSANFPVSWREMPIAEAKAAGATALFGEKYGDVVRVVRVEGDVDFGGHA
                     VSSMELCGGAHVSRTGDIGAFVIVSDENVAAGVRRIEALAGEQATAWLRGRLNASAKA
                     AALLNTSPEGLEERVSGLQGQLKAAEKETAQARRQLAEAQMGGGGSAAQTRELGGFKV
                     ASLKLSGIEGNELRGAADKLLDQSGADLVVIASDKGLVVKATKDAVAKGAHAGQLVSK
                     LAAAGGGKGGGRPDMAQAGITDAAGALGALDTAF"
     gene            2782. .3426
                     /gene="DR2301"
     CDS             2782. .3426
                     /gene="DR2301"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11857.1"
                     /db_xref="GI:6460120"
                     /translation="MIEKGWGVGSSFFVEGVVEEFLSQLDDLDDAVSSVSLSAAQLTG
                     DNLHLTFRLAGQTWDGTDITQSWQVTAQGVAEHRLSLGECGRWPELLSASPRLLVAQA
                     PWRKLMFSTAPADAAAFLSDLRAVHDDEKWDTFGKYGLGGQLDLGYGVFADGPAPLLE
                     LYHPVLEAHRMRPSWLPSAPRPTPPRKCLHLGGTVWKDDLYMLAQAFVATQQPL"
     gene            3460. .3948
                     /gene="DR2302"
     CDS             3460. .3948
                     /gene="DR2302"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11853.1"
                     /db_xref="GI:6460116"
                     /translation="MSETQHWLDGIVDILREALEGGEPGQGTAFVENTGADGVRYGLL
                     PTLARLSAAQASQEVYGTSVAGHARHTAFHLEVLIRWDRDGERGPFDWKGSFLPMQTS
                     EEEWPAVQQRLRDAYEAVVAFEASQRSQAPNGDLSGSFAGAAAHAAYHLGAIRQLIKV
                     VA"
```

```
gene            complement(3945. .4631)
                /gene="DR2303"
CDS             complement(3945. .4631)
                /gene="DR2303"
                /note="similar to GB:L02937 SP:P26826 GB:M74769 PID:144741
                PID:144780 percent identity: 62.56; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="chloramphenicol acetyltransferase"
                /protein_id="AAF11849.1"
                /db_xref="GI:6460112"
                /translation="MFFYHRPRGRIFYYPHHQKVNRYPQPMIFCPFPADWPRREAYAH
                YRRNPCSYSVTQRVDVTALRRSGRPFAPALIHAVCRTVNRLPALRMTEQGGVPGYWDE
                VCPAYTVFHPSTGTFSVLWTDYSPDFADFLTQYRADVERYGDDPAFFPKGPAPAHVVN
                LSIAPWLDFAAFNLDYGRDYLLPVFTAGRFSEEAGQVTLPLAVQANHAACDGWHVAEF
                FRVLGKELER"
gene            complement(4732. .5922)
                /gene="DR2304"
CDS             complement(4732. .5922)
                /gene="DR2304"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF11854.1"
                /db_xref="GI:6460117"
                /translation="MTTVRRTLDVVTPLPLLLAAQCLGVMGSAAYSVAALWTLKAGGG
                DDRLLALAGALFALPVLLLGVPAGRLADRVDRKGILILAPLACAGVLGGVVWLAEGTA
                AALTVPLLLAGVLLNVALTFDEPAFLASLPELAGEGELGRINALHQLLMSAGSIGGAL
                LGGVLASWGVGAAATVGLFAYLLCFLLVPLTRWPARVVVPTSPGIPRSSLSELWCQLS
                LDLRQLVVLVCAFITFAAPMPALLPSLVHDQLQLGPQALGALEAAFGAGFVLGGGLAL
                LGVRPRGLAGLGLAVLVAGAVRLGLSVAGTLPVSAGLLLLGGSAVAVASIAAMTQLQE
                GAPATSRGAVLGAVSSLLAGAAPLGLGLATATLPALGLSGVLQGTGVVLLLLGLGALL
                RGRR"
gene            5952. .6653
                /gene="DR2305"
CDS             5952. .6653
                /gene="DR2305"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF11855.1"
                /db_xref="GI:6460118"
                /translation="MSDSFSTSSPAPACSSECECLTVTDARQAAAMIHPEFSPLMDAL
                TGREASVGALARELDWPLDTAHYRVKKLLRLGLLRVVREEKRAGRPVKVYRAVADAFF
                LPYSSFTHATLEEKFLRHELQVARSLVRSYVRLLHDTFPQPGTLGPYLFRDEAGHLQE
                DLAAAPGKLLDLAAAFPTTTEHLRSLTLTDAEARHLRDVLDALAAEFPRRRAGQPLTH
                MLRLWLVPAREEDWQ"
gene            complement(6766. .7233)
                /gene="DR2306"
CDS             complement(6766. .7233)
                /gene="DR2306"
                /note="similar to SP:P22853 PID:1129093 percent identity:
                63.81; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, MerR family"
                /protein_id="AAF11850.1"
                /db_xref="GI:6460113"
                /translation="MTIMTEPARYKVGEVAARLGLTLRTLKYYEELGLVTPERTDSRY
                RMYSEADVQRLERIRRMRALGLSLGTIQAISTHPQESDPDGRQVLTQEALHTLESDLS
                SQLTVLSTRITAAEKELKDARALRRDVQRDLDYVQRRLAGAQVGELLREKDPA"
gene            complement(7445. .8677)
                /gene="DR2307"
CDS             complement(7445. .8677)
                /gene="DR2307"
                /note="similar to GB:AL009126 percent identity: 64.29;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="multidrug-efflux transporter, putative"
                /protein_id="AAF11851.1"
                /db_xref="GI:6460114"
                /translation="MYVWVMSASASSLPGEVALDPSEASRRNVASWAVTAACLIAFMG
                VGVVDPILPEIGRQLGATPTQVELLFTSYLGVMAVMTLFAGNIGAKLGRKRVALIGLG
                LIALFAAACGLSGSIPALAVFRGGWGLGSALFTPTALVLLLALIGHAEKAIMRYEAAI
                GLGMSMGPLLGGVLGSHSWRFPFFGAATLMLLALASVAAFVKVPESKEPPRPIGDVFR
                AYTRPAFLAVGLTGLLYYFGFFLLLGYTPLFLHLSTLNLGLMFFGWGILLGLGSTKLA
                EVLLRRVKASNVVMGALAGLTIIFALLGFAPFAVGVKVGLVVLSGVLFGLMNATLTTL
                SVEVSHMPRPTATSAYNFLRWLGAAAAPVSSGLIAEHLGGKMPYGFGAGAVALGLVIT
                ALAARQIDAARASDPHVH"
gene            8849. .10165
                /gene="DR2308"
CDS             8849. .10165
                /gene="DR2308"
                /note="similar to PID:1146219 SP:P50743 GB:AL009126
                percent identity: 64.07; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF11852.1"
                /db_xref="GI:6460115"
                /translation="MHKVAIVGRPNVGKSSLFNRLIGRREAVVADFPGVTRDAKEGLM
                LYHNHRITLIDTGGLWSGDEWEQAIREKAEWAMEGAQAVIFVLDPREGLSAADYEVAD
                WLRRLGKPVIVTANKIDSQKHEPYLAELWGLGFGDPVAISAEHARGLDDLMDRVMKYL
                PEDDEDVPDIAPIRISLIGRPNVGKSSLLNAITQSDRAIVADLPGTTRDSLDVEWDYG
                GQRFVLVDTAGIRKKPDTAIEDFAIQRSQAAIERSDVIWLVVNATDIGDHELKLANLA
                YDSGKPVIVVVNKWDLVPDAELKSTEKDLNQKLHHISFAPRVYTSAINDYGIHDMLAE
                AMKLYEKWQSRIGTSELNRWLEVWQMKQRVPNFHGKPLRMYFMTQVETAPPTFAIFCN
                RADFVTRAYEGFLQNRIREDLELAGVPVRLKWKEKGPYKKGEADED"
gene            complement(10225. .10881)
                /gene="DR2309"
CDS             complement(10225. .10881)
                /gene="DR2309"
                /note="identified by Glimmer2; putative"

  Query Match             79.1%;  Score 17.4;  DB 1;  Length 10968;
  Best Local Similarity   94.7%;  Pred. No. 1.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 gcggcgtaggccgggggtc 22
           ||||||||||||||||||| |
Db    3334 GCGGCGTAGGCCGGGGGGC 3316


RESULT   5
AY028171
LOCUS       AY028171               44273 bp    DNA     linear   INV 03-AUG-2001
DEFINITION  Leishmania donovani extra chromosome XC290, partial sequence.
ACCESSION   AY028171
VERSION     AY028171.1  GI:13518079
KEYWORDS    .
SOURCE      Leishmania donovani.
  ORGANISM  Leishmania donovani
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 44273)
  AUTHORS   Dubessay,P., Ravel,C., Bastien,P., Lignon,M.F., Ullman,B., Pages,M.
            and Blaineau,C.
  TITLE     Effect of large targeted deletions on the mitotic stability of an
            extra chromosome mediating drug resistance in Leishmania
  JOURNAL   Nucleic Acids Res. 29 (15), 3231-3240 (2001)
  MEDLINE   21363922
   PUBMED   11470881
REFERENCE   2  (bases 1 to 44273)
  AUTHORS   Dubessay,P., Ravel,C., Bastien,P., Lignon,M.-F., Ullman,B.,
            Pages,M. and Blaineau,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-2001) CNRS UMR5093, Laboratoire 'Genome des
            Parasites', 163, rue Broussonnet, Montpellier F34090, France
FEATURES             Location/Qualifiers
     source          1. .44273
                     /organism="Leishmania donovani"
```

/strain="1S"
/db_xref="taxon:5661"
/chromosome="XC290"
/note="line MPA100
extra chromosome derived from chromosome 19"

CDS 2157. .3794
/note="similar to Oryza sativa phosphatidic acid phosphatase"
/codon_start=1
/product="phosphatidic acid phosphatase-like protein"
/protein_id="AAK27378.1"
/db_xref="GI:13518080"
/translation="MDAFSLPLFPSAATFAHPRFSVLLCTVWPLRTHLHTHTHTHTHR
VDTLHSAKSTPSRTPSQFFKVLCIAHVRPFLSLLVLFVYSRALSPPPAGGCGLPTKEL
FALHLPLQAFVHSCDGSLPYVPTRASHRRSPSSSSHRSPLLFAMVVLSKKEFYYAWNQ
YHLLDWIVLTVLLLISMIVTISMKPHCRTFSWNDATIAYPSHADTFPDYSLALMLVFS
VVFYVIFIWYLVRPLQEFVGEPLDWYSIGGADAGTSGEGNAYAEMDTVVNTRQPRIRD
MQTGRGLVYPWLRAQLWSVGMESCATAVLKVYAGRLRPDYLSRLKAAGYTSSMSHLPD
PQTNPDYYCALMDAHPALKEGRLSFPSGHSSTSFAVCTVVCLFFVAHLRPFARHASFT
RLIICLLPISVSLMCAVSRTRDNKHHFSDIIAGSLIGVVSAFLSFYGSFRQVGGAAGI
YFSRTATDIEYEQLREWKSAGGNEGSGAAVEIPTAGGSSANYQSTNKREEAARCRSAS
PINRMNDAVVTFNDSVRRPTGLTERRLNEDPAAVPWI"

repeat_region 6379. .7081
/rpt_family="IR2"
/rpt_type=dispersed

repeat_region 7971. .8160
/rpt_family="IR1"
/rpt_type=dispersed

CDS 9756. .11999
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AAK27379.1"
/db_xref="GI:13518081"
/translation="MASSTELMPHSERDGSEGSVPYPHRVGETLVLELNTLCRLRDFY
EQHTFDTQGPYADLFAVLLSSCKHVIDINVEFVHHIKAQQDQQRRQQVACDVLCREAA
LQRADIKHLRAALEHLSARRASGEASVVKCEDDDTGLGNGSAASSHASPKNEAATAEC
CQHVSQRSPAASLAPQRSTGAASPSSAAVGLERSSLSTASASLVGPVSQAHATVTFAP
TEAPAQPINSLTHSSPSDHACGGAWAAEMVRRTDAMQGSSATVLVDGQLSSVRLDEEQ
RKQQQSSTALFAPSEARKHHAEDAVESHLEHLTRHFAAVDPRVHKQRRMWGEEEDVGS
RSPSSSPVAASGTAGESEFPHSCTTVAGHARSNPQESSQRQGTSVASPVGTWTPPAVL
VRWLDSRLRRWKAEWKRVLEDVQDALLLDATVDAQTLRPIAHAAGQQEAHVNLGTAHE
RASESNGGAEAVMTPSCRGDLVCVIRALLSSHGTELYKQIVAESTRQVQRTDEELHDL
RSRLEALEVYAPHRYAVTARPPLLGVELEDVLEPRIGVRLRTVYHHYLADRAGLSVGD
VLVGVGHQSIQTRAQLHVVLGELTRDYNAQCHIQIEEGFMRSFALGDQCCVRDGHCVE
CHSPYREATVERKGNDYGLGQELQRARSEAAAAVAGGTIPSDARSSARWMYMHSSLKG
LAAMAGSSAAASPFLSTFHSAAQHKEKLAQYLPYFELCLHVVRDGRLRDVTLLIPPWE
ALRSAAY"

repeat_region 14182. .14759
/rpt_family="IR1"
/rpt_type=dispersed

CDS 15348. .17207
/note="similar to Saccharomyces cerevisiae mitochondrial metal transporter"
/codon_start=1
/product="mitochondrial metal transporter-like protein"
/protein_id="AAK27380.1"
/db_xref="GI:13518082"
/translation="MRRFFAMKPSAVAKTTTAAGTAAQLHAHSHRHGGHAHAHSHGSI
EKGISGKYLRQCQVATLAGGATNVFFCVTKLWIGSAGGSVALVADGFHSLTDILADII
SYVAISLSRKKLPRCRFPLGIGRLETSGAVIVAAILFFGGVALLVQSLEQCFADVASL
LETAAPTGGIAAALALGARRAMAALRGNSTSAGSPAGHVHAHGSHDHSHEENHGGGDH
QGHSHFQVMTYDEDIGKQVILWTMVGVATASVVCKELLFRWTRRVGLRAGSRVVVANA
YHHRADAWSGGVSLLGVAGQVMGLPGVDGLAGLAVSLSICKIGYGIFKDSVLEFFDYQ
NAEEVSAIRDQLQRFNLCIVRGVDVVPNHKCQNAVKAAASGTEASSTSSPATASSSDD
DETVHRKKMHFINVFLVRHGHQYAVHVTLLVYESVSAQQIKEAADLITALARRSLPVQ
DTFTTLLVCSNYTESREVSLGGCNDNEGVDEEGAAAISSGATAANCHGHSRAGCSGSF
SRAHGSEDSATSTPPIAATGAIINPSLERCIASLQEFHTFTAPIRYSWEERTITAPSS
ESCECERDINSVAEMFKCRLISGEIASAKKTAADSAKRSGHSCSHGAQADHSH"

CDS 17975. .19162
/note="similar to Schizosaccharomyces pombe transcription initiation factor"
/codon_start=1
/product="transcription initiation factor-like protein"
/protein_id="AAK27381.1"
/db_xref="GI:13518083"
/translation="MCVCVCGDAQVLSLHVFLCFSPPSHPAQTLLSLPCIPCSLELEN
SLLRLFINHSLLLKPVCLSLRGGAHSSVVPHPSLLCSALIRALPCARLIFDPFLCRST
MDDDYAFFESDNLVEEELPVIGAFPDPVAEFGGVVGPTGAIGDSALTADGHLAAAAAT
LPGGDDAAANLDVDELPPPIKNVQEFLPNVHPDAFPVVVAVQAQASIPVGINLAELSC
ATRNVEYMPNNRIPPATMRLHEPTAVVMMHNSGALSIIGAASVSEARQAAELAARIIR
KALNLNFSSLKFRVRSIAARFNVCSPIRLDKLAAYQLDPAMSIGVAKLQVSYEPERFN
GCVLRLVGKSSRGDNQWSVSCSVFVTGKVQLMGARSMDELRFAFNAFVPIIAKNLDER
KTA"

CDS 19641. .20714
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AAK27382.1"
/db_xref="GI:13518084"
/translation="MQQHSFTSLVDSQTSQAPRSLFLLNASMTVQRHHSGARAHGKKE
IDWHGCLSSLPLSAPFSALLSLLQNLSWESLHVHKRVHIRTRLTLPTHPPTPSTSIII
ATTPRGHLTHPTQHLSAFSFFILLYITMSDWETIAKESEEVLAKPDIKGCHEILMKAY
EGGNQHPEILWRLGRSYYEMANESTDPAVREPYLKEGMELCKKSVEADPNNFASHKWL
GILISDQKVGNKEKIANAYVIRDHFLKAVELNPNDATSLHCMGNWCFKILQIGWLERK
AAALILGEPPSSTYEECLGYLLRSAEAGNTIHNSTMIGDAYAQQSMHDEARKWYQKAI
DMPAYTELQKRNHDVAVAKMKKL"

repeat_region 21526. .22103
/rpt_family="IR1"
/rpt_type=dispersed

CDS 22929. .25730
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AAK27383.1"
/db_xref="GI:13518085"
/translation="MTEKEANDLIASAFSSEAMVEAVARVVASGYEKCSLCNNIIEGE
AEHVLREHPHVQHIHTMVAKPISPLPARGSQDQASMDVGSFAVATSVKADRGEDAFRV
VQRARAQLHRLLDGMQGIEGHVYLFGSVVSMGSWDGVGDGDFSFICPKWYTLPAPHGQ
VVVREEAVEEDSDDETLAGRKEGKEVEESSSVGEADDFGTVSATALRHASPGKSAEDA
ATEKSQSSVLSLKKEKRIICNIAARLRDAGFRFEELDLVLSTRIPVVRRKRAEQEPLR
LHAAPQGSFICVRFDTAHAEEQFRNERLSRVISTYNAEEVPSRGNQRGRSLVLCVPDS
TDAIHLMSQWKRKNGVRREWLGNRRSPEIFSIDFDLSCRCNGIRNSWLLRRYLAQSEV
FRVGNVFLKTWSKACGVNNSRVGFLTSYAVSVLWIYFLLRRGEVVFVNPADIPALPNP
KEQMEVPYIPLWPAVADAEADAARTARLGSLLRGFFYFYGEEFDWGAHVATIRQPCDS
AADIRTKEDLGWEKSDTLSLVLRDRCYHIFSIEDVYEDDLDLGRHLTPEKAGWIRLQF
RLAYRLCCSTRRAGGAAGPSPLWQLFDTPRKRAEDVLHARLYAYLLRNTEDAAAPISE
ILEQLGVDEESEAVDADTKEDPLYLACAYELGNRLCDLWFDDAQVAQDTAFHKMYNRR
SMDYTPPANPTTCGEWAAVCTDDLQEHHDPTSVHQQRCVRLVPPLRSEAGAPAPHFQA
ANAPARERALQLIKSRRQYEECARELGAHAAKKSNQAPAAVVGVPLKKSAAGMLYPHC
FYPLQQHRLFETYAARATFTRCVRDVIEEMALLRGADAPAAFGERGDVLHDRNKLLRY
LKKRLCSSVVVADPTYEAVLQFVCLPTEPYIHAPQLPHGHNERPTLHPTPLFLSLVGD
GSSGAGVAAKAKVPRARPP"

CDS 26821. .27885
/codon_start=1
/product="cysteine proteinase-like protein"
/protein_id="AAK27384.1"
/db_xref="GI:13518086"
/translation="MARRNPFFFAIVVTILFVVCYGSALIAQTPLGVDDFIASAHYGR
FKKRHGKPFGEDAEEGRRFNAFKQNMQTAYFLNAHNPHAHYDVSGKFADLTPQEFAKL
YLNPNYYARHGKDYKEHVHVDDSVRSGVMSVDWREKGVVTPVKNQGMCGSCWAFATTG
NIEGQWALKNHSLVSLSEQVLVSCDNIDDGCNGGLMEQAMQWIINDHNGTVPTEDSYP
YTSAGGTRPPCHDNGTVGAKIAGYMSLPHDEEEIAAYVGKNGPVAVAVDATTWQLYFG
GVVTLCFGLSLNHGVLVVGFNRQAKPPYWIVKNSWGSSWGEKGYIRLAMGSNQCLLKN
YAVTATIDDSNTSHVPTTAA"

repeat_region 28652. .29379
/rpt_family="IR2"
/rpt_type=dispersed

repeat_region 37043. .37481
/rpt_family="IR1"
/rpt_type=dispersed

CDS 37687. .38784
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AAK27385.1"

```
                     /db_xref="GI:13518087"
                     /translation="MLTASGLMVGFAFVAKTADESCSYIVNVCGHDSVGPPLARSMNA
                     VDTEYVEHRGVDNLIIPISVSEPKETKSGDYAYCVDVVVHTLLIKKCSPGHRLLQHLT
                     EKLVTLSMEWIRNECGVQLLPRSCRIAGNPLYYDDTVETSKSMTSIIENAAALLQRKS
                     NSSFTRKDEQDTIPDALDLMKAQVPECKQPRIREVISTPAIKKGFLVDGNARLYGPDG
                     SSEGSGKAPDPLSHIPQSLRERCQIIDTRNVEGLTAPPAKTLKSVDKTANTHTSGTIV
                     APTSELQGKDEASWARFSIERTGEKVIVRFLVPDSVSSLHDIDLSATKDTLEINGSVT
                     QLPVPIVTDDVQAKFVKSTRILIVTCMIDVS"
     CDS             39073. .40164
                     /codon_start=1
                     /product="mitogen-activated protein kinase-like protein"
                     /protein_id="AAK27386.1"
                     /db_xref="GI:13518088"
                     /translation="MAQLVPLVELPSGKKIYSVRGQQFEVDRQYDLVKVVGFGACGTV
                     CSAVANGSGERVAIKRLSRVFGDLREGKRILREMEIMTSLKHNNLIRLHHFMRPQSKE
                     TFEDIYLVMDLYDTDLNRIIRSRQKLTDEHLQYFMIQAFRGLHYLHSAKVMHRDLKPS
                     NLLVNADCALAICDFGLARDDQVMSSSDLTQYVVTRWYRPPEVLGMGSNQYTSAVDVW
                     SLGLIFAELMLGRTLLPGTDYIGQLVMIVNLLGSPSIDDMEFLSSEAKAFILSQPRRP
                     ALSFRDLFPMATEEATDLLSKLLVFHPARRLTAKQVMEHPYFSKYRDAAEEADAPDPF
                     VWNHSHIETKEQLREDLWRVVEAHSQSNG"
     CDS             40379. .42319
                     /codon_start=1
                     /product="oxydoreductase-like protein"
                     /protein_id="AAK27387.1"
                     /db_xref="GI:13518089"
                     /translation="MKRKPILCALLPNLYFLPFSPPLNSTWILRSFFIFYFFFIYFCS
                     SRDNYSTISFVYTGKLRRMRRPREPAPGECCGSGCTRCVWDIYYDEVARFEELIAGGG
                     IEEDCTQSSEEEEVVNYIGSVVVKYIDPPALPTTGSPGEWERAEMKARGFFPIDRIEL

  Query Match           79.1%;  Score 17.4;  DB 3;  Length 44273;
  Best Local Similarity  94.7%;  Pred. No. 1.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 gcggcgtaggccgggggtc 22
           ||||||||| |||||||||
Db     372 GCGGCGTAGTCCGGGGGTC 390


RESULT   6
HSU53365
LOCUS       HSU53365                2317 bp    DNA     linear   BCT 11-JUN-2001
DEFINITION  Halobacterium salinarium transducer HtI protein (HtI) gene,
            complete cds.
ACCESSION   U53365
VERSION     U53365.1  GI:1589744
KEYWORDS    .
SOURCE      Halobacterium salinarum.
  ORGANISM  Halobacterium salinarum
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 2317)
  AUTHORS   Zhang,W., Brooun,A., McCandless,J., Banda,P. and Alam,M.
  TITLE     Signal transduction in the archaeon Halobacterium salinarium is
            processed through three subfamilies of 13 soluble and
            membrane-bound transducer proteins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (10), 4649-4654 (1996)
  MEDLINE   96209786
REFERENCE   2  (bases 1 to 2317)
  AUTHORS   Alam,M., Zhang,W. and Brooun,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1996) Maqsudul Alam, Microbiology, University of
            Hawaii, 2538 The Mall, Snyder Hall #207, Honolulu, HI 96822, USA
FEATURES             Location/Qualifiers
     source          1. .2317
                     /organism="Halobacterium salinarum"
                     /strain="Flx15"
                     /db_xref="taxon:2242"
                     /note="transducer HtA, HtB, HtC, HtD and HtF protein genes
                     have been deposited in GenBank Accession Numbers
                     U75435-U75439, respectively"
     gene            663. .2297
                     /gene="HtI"
     CDS             663. .2297
                     /gene="HtI"
                     /note="methyl-accepting taxis protein"
                     /codon_start=1
                     /transl_table=11
                     /product="transducer HtI protein"
                     /protein_id="AAB17519.1"
                     /db_xref="GI:1621047"
                     /translation="MSGAAVFVDAVVAPLGDAVGAIGFGAAAALGYRNYRDTDAEAAF
                     WMAFTFASLLGVTWTVSLMLEKAGVATQIFNLATGPLMATTVAVFAIGGTATLAIVED
                     MEALVEERAQPRQEAEEERAEAERAREKAEQKQAEAERQTAEAQSQTGRTGTQREIEQ
                     LAADLESQATEVGATLEAASDGDLTARVDATTDNAEIAEVATVVNDMLTTMERTIDEI
                     QGFSTNVTTASREATGGAKEIQVASQTVSESVQEIAAGTDDQREQLESVAEEMDSYSA
                     TVEEVAATAQSVADTAADTTDVATAGKQTAEDAIDAIDAVQETMQTTVANVDALEDLT
                     TEIDDIAELISDIAEQTNMLALNANIEAARAGSGGGTNGDGFAVVADEVKELATESQR
                     SAKDIAELIEEVQSQTATTVEEIRVAEQRVNDGAAAVEETVDAFGAVTENIQETTDGV
                     QEISQAMDEQAQRSERVVSSVDDIATISQATADRAENVSAASEEQTASITEVTGSLQS
                     LAAQADTLEDRLNEFRTEATGTAHGERTDAPAGQSD"
BASE COUNT      469 a    721 c    824 g    303 t
ORIGIN


  Query Match           78.2%;  Score 17.2;  DB 1;  Length 2317;
  Best Local Similarity  86.4%;  Pred. No. 3.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           |||||||||| || ||||| ||
Db    1191 GAAGCGGCGTCGGACGGGGATC 1212


RESULT   7
HSHTPIVAV/c
LOCUS       HSHTPIVAV               4287 bp    DNA     linear   BCT 01-AUG-1996
DEFINITION  H.salinarium htpIV and htpV gene.
ACCESSION   X95589
VERSION     X95589.1  GI:1435130
KEYWORDS    htpIV gene; htpV gene; transducer protein.
SOURCE      Halobacterium salinarum.
  ORGANISM  Halobacterium salinarum
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 4287)
  AUTHORS   Rudolph,J., Nordmann,B., Storch,K.F., Gruenberg,H., Rodewald,K. and
            Oesterhelt,D.
  TITLE     A family of halobacterial transducer proteins
  JOURNAL   FEMS Microbiol. Lett. 139 (2-3), 161-168 (1996)
  MEDLINE   96275896
REFERENCE   2  (bases 1 to 4287)
  AUTHORS   Rudolph,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-1996) J. Rudolph, Max-Planck-Institut for
            Biochemistry, Am Klopferspitz 18a, D- 82152 Martinsried, FRG
FEATURES             Location/Qualifiers
     source          1. .4287
                     /organism="Halobacterium salinarum"
                     /strain="S9"
                     /db_xref="taxon:2242"
     gene            101. .2533
                     /gene="htpIV"
     CDS             101. .2533
                     /gene="htpIV"
                     /codon_start=1
                     /transl_table=11
                     /product="halobacterial transducer protein IV"
                     /protein_id="CAA64841.1"
                     /db_xref="GI:1435131"
                     /db_xref="SWISS-PROT:Q48317"
                     /translation="MSEPTADAGDNSPSSTDTGPLDRVKAIALLPLRSYLVKFAVALL
                     VILVIIAAGGFWVQADATATLEANTEQQLEQEAVSDATEIGDWLERNEQSVLIASNNP
                     RLGFNTTAADKQAYVTQLVAAELDADRIADVHVADPTVGGASDARIVASTDEDARGTR
                     VSADTHPWVDRTRSIGRDTVVSTNPYRTAGGQRVVSSMSVAADLTHVLVVEYTAGDLS
                     DQFGAGIDGTFTQVVRPTSDATAVLFSDAGTDAVGQPYIPDRSQSEIPALDSATEQGQ
                     FTNTPTKDSVLDREYVAAYTTVPGKNWVVVKHAPSESAFALSNQIRTGILGFILVALV
```

```
                     GVVLVGGTIGRNTAAAVQSLSAAAAEIEAGNYDVDVASSRRDEIGQLFASIGSMRDAL
                     VTQIDEAEAAREQATEAQQDAEAERERAEDARERAEDAKADAEALAAELEAQAERYSD
                     VMAACADGDLTRRMPADDTDNEAMAAIAASFNEMLAQWEHTIIDIQEFADAVATASEE
                     AEVGAADAERASGQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAASADSVA
                     EHSHQTAEIARDGEQTAEDAIERSLSVQEAIDATVQNVEALDDQMAEISEIVDLISDI
                     AEQTNMLALNANIEAARADKSGDGFAVVADEVKDLAEETQESAGDIERRITEVQSQTT
                     ATVAEARAAEESMDAGIDAVEEVVDAFTAVSDHADETDTGVQEISDTTDDQASSTEEA
                     VSMTEEVADLSDSTAGEAQSVSAAAEEQAASMSEISDSVESLSGQAEQLKALLSEFEV
                     DADRDVTPTQTD"
     gene            complement(2549. .4186)
                     /gene="htpV"
     CDS             complement(2549. .4186)
                     /gene="htpV"
                     /codon_start=1
                     /transl_table=11
                     /product="halobacterial transducer protein V"
                     /protein_id="CAA64842.1"
                     /db_xref="GI:1435132"
                     /db_xref="SWISS-PROT:Q48318"
                     /translation="MSGAAVFVDAVVAPLGDAVGAIGFGAAAALGYRNYRDTDAEAAF
                     WMAFTFASLLGVTWTVSLMLEKAGVATQIFNLATGPLMATTVAVFAIGGTATLAIVED
                     MEALVEERAQRRQEAEEERAEAERAREKAEQKQAEAERQTAEAESAKQDARERSAEIE
                     QLAADLESQATEVGATLEAASDGDLTARVDATTDNAEIAEVATVVNDMLTTMERTIDE
                     IQGFSTNVTTASREATAGAKEIQDASQTVSESVQEIAAGTDDQREQLESVAEEMDSYS
                     ATVEEVAATAQSVADTAADTTDVATAGKQTAEDAIDAIDAVQETMQTTVANVDALEDL
                     TTEIDDIAELISDIAEQTNMLALNANIEAARAGSGGGSNGDGFAVVADEVKELATESQ
                     RSAKDIAELIEEVQSQTATTVEEIRVAEQRVNDGAAAVEETVDAFGAVTENIQETTDG
                     VQEISQAMDEQAQRSERVVSSVDDIATISQATADRAENVSAASEEQTASITEVTSSLQ
                     SLAAQADTLEDRLNEFRTEATGTAHGEATDAPAGQSD"
BASE COUNT      675 a   1494 c   1445 g    673 t
ORIGIN


  Query Match           78.2%;  Score 17.2;  DB 1;  Length 4287;
  Best Local Similarity   86.4%;  Pred. No. 2.8e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gaagcggcgtaggccgggggtc 22
          |||||||||| || ||||| ||
Db   3655 GAAGCGGCGTCGGACGGGGATC 3634


RESULT   8
DVHB/c
LOCUS       DVHB                    8807 bp    DNA     linear   INV 20-JUN-1994
DEFINITION  Drosophila virilis hunchback (hb) gene for zinc-finger protein
            transcription factor.
ACCESSION   X15359
VERSION     X15359.1  GI:9178
KEYWORDS    developmental regulation; hb gene; hunchback gene; segmentation
            gene; transcription factor; zinc finger protein.
SOURCE      Drosophila virilis.
  ORGANISM  Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 8807)
  AUTHORS   Tautz,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-1989)
REFERENCE   2  (bases 1 to 8807)
  AUTHORS   Treier,M., Pfeifle,C. and Tautz,D.
  TITLE     Comparison of the gap segmentation gene hunchback between
            Drosophila melanogaster and Drosophila virilis reveals novel modes
            of evolutionary change
  JOURNAL   EMBO J. 8 (5), 1517-1525 (1989)
  MEDLINE   89356623
COMMENT     *source: EMBL 4;
            Data kindly reviewed (04-OCT-1989) by Tautz D.
FEATURES             Location/Qualifiers
     source          1. .8807
                     /organism="Drosophila virilis"
                     /db_xref="taxon:7244"
                     /clone="lambda-hbDv-1, lambda-hbDv-5/2"
                     /clone_lib="lambda"
     mRNA            921. .1534
                     /note="P1 transcript, exon 1"
     precursor_RNA   921. .8714
                     /note="zinc-finger protein P1 transcript"
     intron          1535. .5617
                     /note="P1 transcript, intron I"
     mRNA            5246. .5350
                     /note="P2 transcript, exon 1"
     precursor_RNA   5246. .8714
                     /note="zinc-finger protein P2 transcript"
     intron          5351. .5617
                     /note="P2 transcript, intron I"
     mRNA            5618. .8714
                     /note="exon 2"
     CDS             5641. .8091
                     /note="zinc-finger protein (AA 1 to 816)"
                     /codon_start=1
                     /protein_id="CAA33419.1"
                     /db_xref="GI:9179"
                     /db_xref="FLYBASE:FBgn0013116"
                     /db_xref="SWISS-PROT:P13361"
                     /translation="MPNWETSTAAPSYEQHNAWYSSMFAANIKQEPLSHHHHHHGQQH
                     HDNHSNSNSGASSPRQSPLPSPIPPSTQLEQYLKQQQQQQQQQQPMDTLCAAAMTPS
                     PSNNDQNSLQHFDATLQQQLLQQQQYQQHFQAAQHQQQQHHHLALGGFNPLTPPGLPN
                     PMQHYYGGNMRPSPQPTPTAAPTAVAAAIQTGDKLQALTPPMDVTPPKSPAKSQQSSA
                     EPEKEHDLMSNSSEDMKYMAESEDDDSNIRMPIYNSHGKMKNYKCKTCGVVAITKVDF
                     WAHTRTHMKPDKILQCAKCPFVTEFKHHLEYHIRKHKNQKPFQCDKCSYTCVNKSMLN
                     SHRKSHSSVYQYRCADCDYATKYCHSFKLHLRKYGHKPGMVLDEDGTPNPSLVIDVYG
                     TRRGPKSKSFSGSGSSCSSTSKRSNASAAAAQQQQQPVATSQLSAALQGFPMPAAAAG
                     TAAGAAGTAAPAAVAPVSPPSPAKSVASVEQAPSLPSALLPPLASLLQQNRNMAFFPY
                     WNLNLQVLAAQQQAAVLAQLSPRMADQLQQQQQQQHQQQQQQQQQQQQQQQQQQLPAHSE
                     NEEDEEEEEHEDDFERKSVDSAMDLSQGTPVKEEPQQQQQQQQLPHSNHMAINLKLKD
                     EDTPLISSSSASRRKGRVLKLDTLLQLKSAAMSSPEQQLKLPASVLPTASSPIAGSSA
                     NKQLADDPCSGASSADESMETGRVPQVNISASSTASSSGNSSNASSSTSNPTAAATVA
                     TSGTVSSSSSSSTTTSSSAPAIYECKYCDIYFKDAVLYTIHMGYHSCDDVFKCNMCGE
                     KCDGPVGLFVHMARNAHS"
BASE COUNT     2563 a   2110 c   1802 g   2332 t
ORIGIN


  Query Match           78.2%;  Score 17.2;  DB 3;  Length 8807;
  Best Local Similarity   86.4%;  Pred. No. 2.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gaagcggcgtaggccgggggtc 22
          || |||||| | ||||||||||
Db   2451 GAGGCGGCGCACGCCGGGGGTC 2430


RESULT   9
AE005080
LOCUS       AE005080                10784 bp    DNA     linear   BCT 12-FEB-2001
DEFINITION  Halobacterium sp. NRC-1 section 111 of 170 of the complete genome.
ACCESSION   AE005080 AE004437
VERSION     AE005080.1  GI:10581211
KEYWORDS    .
SOURCE      Halobacterium sp. NRC-1.
  ORGANISM  Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 10784)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     From the cover: genome sequence of halobacterium species NRC-1
```

```
  JOURNAL   Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
   PUBMED   11016950
REFERENCE   2  (bases 1 to 10784)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES             Location/Qualifiers
     source          1. .10784
                     /organism="Halobacterium sp. NRC-1"
                     /strain="NRC-1"
                     /db_xref="taxon:64091"
     gene            71. .322
                     /gene="eif1a2"
                     /note="VNG1756G"
     CDS             71. .322
                     /gene="eif1a2"
                     /note="Eif1a2"
                     /codon_start=1
                     /transl_table=11
                     /product="translation initiation factor eIF-1A"
                     /protein_id="AAG19983.1"
                     /db_xref="GI:10581212"
                     /translation="MPNDDEVFAVVKRHDGGNHVTLQCEDGKERMGRIPGRMKYRTWI
                     NEGDVVLAEPWDWQDEKANVEWRYSDQDADQLREEGHIE"
     gene            529. .816
                     /gene="VNG1758H"
     CDS             529. .816
                     /gene="VNG1758H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng1758h"
                     /protein_id="AAG19984.1"
                     /db_xref="GI:10581213"
                     /translation="MAAIGAFGVLEVVGLIVAVLGLFPVVSQYKESTKLFTVGYLLLV
                     VGMVATNLEAVVLPVVLNATEHVVGIGLAGVTFLFAAYQRRQGMTATEDTA"
     gene            813. .2450
                     /gene="htr7"
                     /note="VNG1759G"
     CDS             813. .2450
                     /gene="htr7"
                     /note="Htr7"
                     /codon_start=1
                     /transl_table=11
                     /product="Htr7 transducer"
                     /protein_id="AAG19985.1"
                     /db_xref="GI:10581214"
                     /translation="MSGAAVFVDAVVAPLGDAVGAIGFGAAAALGYRNYRDTDAEAAF
                     WMAFTFASLLGVTWTVSLMLEKAGVATQIFNLATGPLMATTVAVFAIGGTATLAIVED
                     MEALVEERAQRRQEAEEERAEAERAREKAEQKQAEAERQTAEAESAKQDARERSAEIE
                     QLAADLESQATEVGATLEAASDGDLTARVDATTDNAEIAEVATVVNDMLTTMERTIDE
                     IQGFSTNVTTASREATAGAKEIQDASQTVSESVQEIAAGTDDQREQLESVAEEMDSYS
                     ATVEEVAATAQSVADTAADTTDVATAGKQTAEDAIDAIDAVQETMQTTVANVDALEDL
                     TTEIDDIAELISDIAEQTNMLALNANIEAARAGSGGGSNGDGFAVVADEVKELATESQ
                     RSAKDIAELIEEVQSQTATTVEEIRVAEQRVNDGAAAVEETVDAFGAVTENIQETTDG
                     VQEISQAMDEQAQRSERVVSSVDDIATISQATADRAENVSAASEEQTASITEVTSSLQ
                     SLAAQADTLEDRLNEFRTEATGTAHGEATDAPAGQSD"
     gene            complement(2466. .4898)
                     /gene="htr5"
                     /note="VNG1760G"
     CDS             complement(2466. .4898)
                     /gene="htr5"
                     /note="Htr5"
                     /codon_start=1
                     /transl_table=11
                     /product="Htr5 transducer"
                     /protein_id="AAG19986.1"
                     /db_xref="GI:10581215"
                     /translation="MSEPTADAGDNSPSSTDTAPLDRVKAIALLPLRSYLVKFAVALL
                     VILVIIAAGGFWVQADATATLEANTEQQLEQEAVSDATEIGDWLERNEQSVLIASNNP
                     RLGFNTTAADKQAYVTQLVAAELDADRIADVHVADPTVGGASDARIVASTDEDARGTR
                     VSADTHPWVDRTRSIGRDTVVSTNPYRTAGGQRVVSSMSVAADLTHVLVVEYTAGDLS
                     DQFGAGIDGTFTQVVRPTSDATAVLFSDAGTDAVGQPYIPDRSQSEIPALDSATEQGQ
                     FTNTPTKDSVLDREYVAAYTTVPGKNWVVVKHAPSESAFALSNQIRTGILGFILVALV
                     GVVLVGGTIGRNTAAAVQSLSAAAAEIEAGNYDVDVASSRRDEIGQLFASIGSMRDAL
                     VTQIDEAEAAREQATEAQQDAEAERERAEDARERAEDAKADAEALAAELEAQAERYSD
                     VMAACADGDLTRRMPADDTDNEAMAAIAASFNEMLAQWEHTIIDIQEFADAVATASEE
                     AEVGAADAERASGQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAASADSVA
                     EHSHQTAEIARDGEQTAEDAIERSLTVQEAIDATVQNVEALDDQMAEISEIVDLISDI
                     AEQTNMLALNANIEAARADKSGDGFAVVADEVKDLAEETQESAGDIERRITEVQSQTT
                     ATVAEARAAEESMDAGIDAVEEVVDAFTAVSDHADETDTGVQEISDTTDDQAASTEEA
                     VSMTEEVADLSDSTAGEAQSVSAAAEEQAASMSEISDSVESLSGQAEQLKALLSEFEV
                     DADRDVTPTQTD"
     gene            complement(4903. .5907)
                     /gene="proX"
                     /note="VNG1762G"
     CDS             complement(4903. .5907)
                     /gene="proX"
                     /note="ProX"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter"
                     /protein_id="AAG19987.1"
                     /db_xref="GI:10581216"
                     /translation="MMDTPEHASTSSRRQLLGMLAAGGTTAVAGCTTITGGGDTTSTG
                     GGDSGSSSIVVGSKGYGEQVTLGYVAYELLANTTNATLVDETGFGGNAAISEAYRTGN
                     VHAYYDYMGSLWSSHPPKHDSADFDTPDAQYDALKSEMESEHPIRILDRADWQNTWAV
                     FVAERAVEGTGIQSISDLAAHVNSRNYDIRPAFGDGFRSRSDGLDALLDYYEFDPERV
                     TAWASEREFLETASAQAAGTAVDEGYADLGVGYSTSAWLTDVDSVRVLDDDRNFWPFF
                     HPVGVVHEDVATDAVVAALNTMPDAIPDAETMQALNSRAAASGNQQAAADHLQNNDFI
                     "
     gene            complement(7701. .8414)
                     /gene="sop2"
                     /note="VNG1764G"
     CDS             complement(7701. .8414)
                     /gene="sop2"
                     /note="Sop2"
                     /codon_start=1
                     /transl_table=11
                     /product="sensory rhodopsin II"
                     /protein_id="AAG19988.1"
                     /db_xref="GI:10581217"
                     /translation="MALTTWFWVGAVGMLAGTVLPIRDCIRHPSHRRYDLVLAGITGL
                     AAIAYTTMGLGITATTVGDRTVYLARYIDWLVTTPLIVLYLAMLARPGHRTSAWLLAA
                     DVFVIAAGIAAALTTGVQRWLFFAVGAAGYAALLYGLLGTLPRALGDDPRVRSLFVTL
                     RNITVVLWTLYPVVWLLSPAGIGILQTEMYTIVVVYLDFISKVAFVAFAVLGADAVSR
                     LVAADAAAPATAEPTPDGD"
     gene            complement(8416. .10710)
                     /gene="htr2"
                     /note="VNG1765G"
     CDS             complement(8416. .10710)
                     /gene="htr2"
                     /note="Htr2"
                     /codon_start=1
                     /transl_table=11
                     /product="Htr2 transducer"
                     /protein_id="AAG19989.1"
                     /db_xref="GI:10581218"
                     /translation="MGSGLVARIRGSYGTKLTLALVVVVVLSVGVGTFVYQQTTTQLE
                     TDVRADLTGSADARADHLDAWLSNARGQTQLASRHPVLASGNDTAITRYLEGLAASDE
                     RPDGVVAAHVYNTSTTTIEASSADAFTGVNPREQGAPFATDPPSFATTSDVVVAAPFT
                     VPAADFPVLSVLSPIPGTTDKALIYMVNVNTLTDDFGQNVAGSTTVVSADGTYVSHP
                     DQDRVLTGHDGPSRLLNQSRTQPAYIDANGTVTAAAPVDGAPWSVLVRAPHDRAFALG
                     DFVASSLVGLVLITIVSLSLIGVTVGSTTVTALRQFSRRADEMAAGDLDTDIDTSRND
                     EFGTLAESFRSMRDSLSESLTDAERATARAEDAREDAEQQRADAEAAREDAEAARKDA
                     QETARALESAAADYEEALTAVADGDLTRRVDASRDHDAMARIGHALNDMLDDIETSVA
                     AATAFSDHVSDAAQRVEADAGDAIDAGTDVSTAVDEISDGATEQTDRLHEVAGEVDDL
```

SASAEEVAETVASLADTAGQAASAVDDGRQATEDAVETMDDVADDAEAAADAMDALDS
EMADIGEIVDVIADIADQTNMLALNASIEAARTGADGDGFAVVADEVKTLAEESRDAA
EDIESRLLALQGQVSDVADEMRATSDTVSDGRATVGDAATALDDVVSFVADTDTAAGE
IRAATDRQAHAASRVASAVDEVAGISQETAAQATAVADSAATQTDTLSSVDDAAADLA
DRAAALDDLLAEFDAHDDTEPEDY"

BASE COUNT 1828 a 3332 c 3751 g 1873 t
ORIGIN

Query Match 78.2%; Score 17.2; DB 1; Length 10784;
Best Local Similarity 86.4%; Pred. No. 2.2e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaagcggcgtaggccgggggtc 22
||||||||||| || ||||| ||
Db 1344 GAAGCGGCGTCGGACGGGGATC 1365

RESULT 10
AC094368/c
LOCUS AC094368 28969 bp DNA linear HTG 12-JAN-2002
DEFINITION Rattus norvegicus clone CH230-3O2, *** SEQUENCING IN PROGRESS ***, 21 unordered pieces.
ACCESSION AC094368
VERSION AC094368.4 GI:17941091
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 28969)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 28969)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 20, 2001 this sequence version replaced gi:17062112.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GALX
Center clone name: CH230-3O2
---------- Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 27105 bases at least Q40
Consensus quality: 31177 bases at least Q30
Consensus quality: 34825 bases at least Q20
Estimated insert size: 13703; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1014: contig of 1014 bp in length
*	1015	1114: gap of unknown length
*	1115	2753: contig of 1639 bp in length
*	2754	2853: gap of unknown length
*	2854	4035: contig of 1182 bp in length
*	4036	4135: gap of unknown length
*	4136	5377: contig of 1242 bp in length
*	5378	5477: gap of unknown length
*	5478	6838: contig of 1361 bp in length
*	6839	6938: gap of unknown length
*	6939	8728: contig of 1790 bp in length
*	8729	8828: gap of unknown length
*	8829	10171: contig of 1343 bp in length
*	10172	10271: gap of unknown length
*	10272	11358: contig of 1087 bp in length
*	11359	11458: gap of unknown length
*	11459	12785: contig of 1327 bp in length
*	12786	12885: gap of unknown length
*	12886	14183: contig of 1298 bp in length
*	14184	14283: gap of unknown length
*	14284	15648: contig of 1365 bp in length
*	15649	15748: gap of unknown length
*	15749	17191: contig of 1443 bp in length
*	17192	17291: gap of unknown length
*	17292	18347: contig of 1056 bp in length
*	18348	18447: gap of unknown length
*	18448	19626: contig of 1179 bp in length
*	19627	19726: gap of unknown length
*	19727	20879: contig of 1153 bp in length
*	20880	20979: gap of unknown length
*	20980	22397: contig of 1418 bp in length
*	22398	22497: gap of unknown length
*	22498	23957: contig of 1460 bp in length
*	23958	24057: gap of unknown length
*	24058	25083: contig of 1026 bp in length
*	25084	25183: gap of unknown length
*	25184	26297: contig of 1114 bp in length
*	26298	26397: gap of unknown length

```
            *    26398    27597: contig of 1200 bp in length
            *    27598    27697: gap of unknown length
            *    27698    28969: contig of 1272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .28969
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-302"
BASE COUNT     7129 a    6343 c    6537 g    6958 t    2002 others
ORIGIN
```

Query Match 78.2%; Score 17.2; DB 2; Length 28969;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 gaagcggcgtaggccgggggtc 22
           ||||||||| |||||||| | |
Db   22262 GAAGCGGCGGAGGCCGGGAGGC 22241
```

```
RESULT  11
MSGB27CS
LOCUS       MSGB27CS                38793 bp    DNA     linear   BCT 17-DEC-2001
DEFINITION  Mycobacterium leprae cosmid B27 DNA sequence.
ACCESSION   L78817
VERSION     L78817.1  GI:1377777
KEYWORDS    .
SOURCE      Mycobacterium leprae (clone: cosmid B27) (tissue library: Lorist 6)
            DNA.
  ORGANISM  Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (sites)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE     Use of an ordered cosmid library to deduce the genomic organization
            of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2  (bases 1 to 38793)
  AUTHORS   Smith,D.R., Richterich,P., Rubenfield,M., Rice,P.W., Butler,C.,
            Lee,H.M., Kirst,S., Gundersen,K., Abendschan,K., Xu,Q., Chung,M.,
            Deloughery,C., Aldredge,T., Maher,J., Lundstrom,R., Tulig,C.,
            Falls,K., Imrich,J., Torrey,D., Engelstein,M., Breton,G., Madan,D.,
            Nietupski,R., Seitz,B., Mao,J.I. et al.
  TITLE     Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome
  JOURNAL   Genome Res. 7 (8), 802-819 (1997)
  MEDLINE   97413161
   PUBMED   9267804
COMMENT     This sequence was generated by the Genome Sequencing Center at
            Genome Therapeutics Corporation (Collaborative Research Division),
            100
            Beaver St., Waltham, MA, 02154.  Please contact Doug Smith
            (smith@cric,com) for further information.  The sequence represents
            the
            insert of a Lorist 6 cosmid clone from a mapped set of clones
            constructed from M. leprae genomic DNA isolated from armadillo
            liver
            [3].  The sequence may not represent the entire cloned insert of
            the
            cosmid if an overlapping region was previously sequenced from
            another
            clone.  Coding sequences larger than 60 amino acids were predicted
            on
            the basis of codon usage and homology information.  An attempt was
            made
            to locate the most probable start site based on codon usage,
            homology,
            the presence of a Shine-Dalgarno sequence, or overlapping orf that
            suggested translational coupling.  It is possible that the actual
            start
            site differs from the one selected.
```

```
FEATURES             Location/Qualifiers
     source          1. .38793
                     /organism="Mycobacterium leprae"
                     /specific_host="Dasypus novemcinctus"
                     /db_xref="taxon:1769"
                     /clone="cosmid B27"
                     /tissue_lib="Lorist 6"
                     /note="The liver of the armadillo was used to isolate the
                     Mycobacterium leprae."
BASE COUNT     7607 a   10325 c   11868 g    8993 t
ORIGIN
```

Query Match 78.2%; Score 17.2; DB 1; Length 38793;
Best Local Similarity 86.4%; Pred. No. 1.6e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 gaagcggcgtaggccgggggtc 22
           | |||||||| |||||| ||||
Db   10830 GCAGCGGCGTTGGCCGGCGGTC 10851
```

```
RESULT  12
AC015511/c
LOCUS       AC015511                64189 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-21O11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC015511
VERSION     AC015511.2  GI:9124022
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 64189)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21O11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 64189)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6437606.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L4071
                Center clone name: 21_O_11
            ----------.
            * NOTE: This record contains 63 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
```

```
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      915: contig of 915 bp in length
*      916 1015: gap of      100 bp
*     1016     1923: contig of 908 bp in length
*     1924 2023: gap of      100 bp
*     2024     2898: contig of 875 bp in length
*     2899 2998: gap of      100 bp
*     2999     3917: contig of 919 bp in length
*     3918 4017: gap of      100 bp
*     4018     4951: contig of 934 bp in length
*     4952 5051: gap of      100 bp
*     5052     5959: contig of 908 bp in length
*     5960 6059: gap of      100 bp
*     6060     6991: contig of 932 bp in length
*     6992 7091: gap of      100 bp
*     7092     8030: contig of 939 bp in length
*     8031 8130: gap of      100 bp
*     8131     9064: contig of 934 bp in length
*     9065 9164: gap of      100 bp
*     9165    10075: contig of 911 bp in length
*    10076 10175: gap of      100 bp
*    10176    11119: contig of 944 bp in length
*    11120 11219: gap of      100 bp
*    11220    12156: contig of 937 bp in length
*    12157 12256: gap of      100 bp
*    12257    13093: contig of 837 bp in length
*    13094 13193: gap of      100 bp
*    13194    14112: contig of 919 bp in length
*    14113 14212: gap of      100 bp
*    14213    15190: contig of 978 bp in length
*    15191 15290: gap of      100 bp
*    15291    16236: contig of 946 bp in length
*    16237 16336: gap of      100 bp
*    16337    17271: contig of 935 bp in length
*    17272 17371: gap of      100 bp
*    17372    18328: contig of 957 bp in length
*    18329 18428: gap of      100 bp
*    18429    19340: contig of 912 bp in length
*    19341 19440: gap of      100 bp
*    19441    20352: contig of 912 bp in length
*    20353 20452: gap of      100 bp
*    20453    21343: contig of 891 bp in length
*    21344 21443: gap of      100 bp
*    21444    22366: contig of 923 bp in length
*    22367 22466: gap of      100 bp
*    22467    23352: contig of 886 bp in length
*    23353 23452: gap of      100 bp
*    23453    24416: contig of 964 bp in length
*    24417 24516: gap of      100 bp
*    24517    25431: contig of 915 bp in length
*    25432 25531: gap of      100 bp
*    25532    26459: contig of 928 bp in length
*    26460 26559: gap of      100 bp
*    26560    27513: contig of 954 bp in length
*    27514 27613: gap of      100 bp
*    27614    28558: contig of 945 bp in length
*    28559 28658: gap of      100 bp
*    28659    29536: contig of 878 bp in length
*    29537 29636: gap of      100 bp
*    29637    30502: contig of 866 bp in length
*    30503 30602: gap of      100 bp
*    30603    31423: contig of 821 bp in length
*    31424 31523: gap of      100 bp
*    31524    32488: contig of 965 bp in length
*    32489 32588: gap of      100 bp
*    32589    33539: contig of 951 bp in length
*    33540 33639: gap of      100 bp
*    33640    34604: contig of 965 bp in length
*    34605 34704: gap of      100 bp
*    34705    35624: contig of 920 bp in length
*    35625 35724: gap of      100 bp
*    35725    36641: contig of 917 bp in length
*    36642 36741: gap of      100 bp
*    36742    37685: contig of 944 bp in length
*    37686 37785: gap of      100 bp
*    37786    38719: contig of 934 bp in length
*    38720 38819: gap of      100 bp
*    38820    39740: contig of 921 bp in length
*    39741 39840: gap of      100 bp
*    39841    40703: contig of 863 bp in length
*    40704 40803: gap of      100 bp
*    40804    41686: contig of 883 bp in length
*    41687 41786: gap of      100 bp
*    41787    42708: contig of 922 bp in length
*    42709 42808: gap of      100 bp
*    42809    43771: contig of 963 bp in length
*    43772 43871: gap of      100 bp
*    43872    44872: contig of 1001 bp in length
*    44873 44972: gap of      100 bp
*    44973    45900: contig of 928 bp in length
*    45901 46000: gap of      100 bp
*    46001    46942: contig of 942 bp in length
*    46943 47042: gap of      100 bp
*    47043    47798: contig of 756 bp in length
*    47799 47898: gap of      100 bp
*    47899    48841: contig of 943 bp in length
*    48842 48941: gap of      100 bp
*    48942    49842: contig of 901 bp in length
*    49843 49942: gap of      100 bp
*    49943    50867: contig of 925 bp in length
*    50868 50967: gap of      100 bp
*    50968    52013: contig of 1046 bp in length
*    52014 52113: gap of      100 bp
*    52114    52948: contig of 835 bp in length
*    52949 53048: gap of      100 bp
*    53049    53947: contig of 899 bp in length
*    53948 54047: gap of      100 bp
*    54048    54918: contig of 871 bp in length
*    54919 55018: gap of      100 bp
*    55019    55937: contig of 919 bp in length
*    55938 56037: gap of      100 bp
*    56038    56945: contig of 908 bp in length
*    56946 57045: gap of      100 bp
*    57046    57989: contig of 944 bp in length
*    57990 58089: gap of      100 bp
*    58090    59039: contig of 950 bp in length
*    59040 59139: gap of      100 bp
*    59140    60070: contig of 931 bp in length
*    60071 60170: gap of      100 bp
*    60171    61112: contig of 942 bp in length
*    61113 61212: gap of      100 bp
*    61213    62165: contig of 953 bp in length
*    62166 62265: gap of      100 bp
*    62266    63175: contig of 910 bp in length
*    63176 63275: gap of      100 bp
*    63276    64189: contig of 914 bp in length.
FEATURES             Location/Qualifiers
     source          1. .64189
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-21011"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    12967 a  14096 c  13532 g  16748 t   6846 others
ORIGIN


  Query Match           78.2%;  Score 17.2;  DB 2;  Length 64189;
  Best Local Similarity  86.4%;  Pred. No. 1.4e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
```

```
          |||| |||| ||||| ||||||
Db  22327 GAAGGGGCGCAGGCCCGGGGTC 22306


RESULT  13
AC017402/c
LOCUS       AC017402               65467 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017402
VERSION     AC017402.1  GI:6553584
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 65467)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210598 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .65467
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    18712 a  14546 c  14336 g  17873 t
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 2;  Length 65467;
  Best Local Similarity   86.4%;  Pred. No. 1.4e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           |||| ||||| |||||||||| |
Db  10339 GAAGGGGCGTGGGCCGGGGGGC 10318


RESULT  14
AP003929
LOCUS       AP003929               100814 bp    DNA     linear   HTG 19-JUL-2001
DEFINITION  Oryza sativa chromosome 7 clone OJ1240_G08, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003929
VERSION     AP003929.1  GI:14915712
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1240_G08.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 100814)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1240_G08
  JOURNAL   Published Only in Database (2001) In press
REFERENCE   2  (bases 1 to 100814)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .100814
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="7"
                     /clone="OJ1240_G08"
BASE COUNT    28110 a  22591 c  22672 g  27341 t    100 others
ORIGIN


  Query Match             78.2%;  Score 17.2;  DB 2;  Length 100814;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           |||||||||| |||||| || |
Db  97427 GAAGCGGCGTCGGCCGGCGGCC 97448


RESULT  15
AC005269
LOCUS       AC005269               104278 bp    DNA     linear   INV 07-JUL-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS00764 (D273) and
            DS00501 (D274)), complete sequence.
ACCESSION   AC005269 AC004571 AC004572
VERSION     AC005269.1  GI:3293205
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in tet from P1 clones DS00764
            (D273) and DS00501 (D274)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 104278)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome, region 21C3-21C7
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 104278)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
```

```
            and relationship to other sequences, please visit our sequence
            archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
            email to drosophila@mhgc.lbl.gov.
            Library locations: 135-8, 21-6.
FEATURES             Location/Qualifiers
     source          1. .104278
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="21C3-21C7"
                     /clone="Pls DS00764 (D273) and DS00501 (D274)"
                     /note="DS00764 (d273) extends from bp 1 to bp 38,674 and
                     DS00501 (d274) extends from bp 20,505 to bp 104,278."
BASE COUNT    28186 a  23704 c  23278 g  29110 t
ORIGIN


  Query Match           78.2%;  Score 17.2;  DB 3;  Length 104278;
  Best Local Similarity  86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           |||| ||||| |||||||||| |
Db   78982 GAAGGGGCGTGGGCCGGGGGGC 79003


Search completed: June 23, 2002, 06:32:29
Job time: 63881 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 15:01:12 ; Search time 652.13 Seconds
(without alignments)
57.921 Million cell updates/sec

Title: US-09-747-514A-4
Perfect score: 22
Sequence: 1 gaagcggcgtaggccgggggtc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	17.2	78.2	2093	22	AAS01678	Human T-type calci
	2	17.2	78.2	5846	22	AAL36954	Human musculoskele
	3	17.2	78.2	5848	22	AAL36955	Human musculoskele
c	4	17.2	78.2	7741	22	AAD04756	Human T-type low v
c	5	17.2	78.2	30393	22	AAK67239	Human immune/haema
c	6	16.2	73.6	664	22	AAS25846	Human cDNA encodin
	7	16.2	73.6	780	22	AAK92231	Human cDNA 5'-end
	8	16.2	73.6	786	22	AAH05059	Human cDNA clone (
	9	16.2	73.6	804	22	AAH05438	Human cDNA clone (
	10	16.2	73.6	865	22	AAK93381	Human cDNA clone r
c	11	16.2	73.6	1419	24	ABA90359	Human polynucleoti
	12	16.2	73.6	1768	22	AAH15691	Human cDNA sequenc
c	13	16.2	73.6	1782	24	ABA90358	Human polynucleoti
	14	16.2	73.6	1790	24	ABL34375	Human immune syste
	15	16.2	73.6	1993	22	AAH15850	Human cDNA sequenc
c	16	16.2	73.6	2079	21	AAC76021	Human ORFX ORF1576
	17	16.2	73.6	2122	22	AAK94529	Human full-length
	18	16.2	73.6	2133	21	AAZ52455	HTRM clone 156986
	19	16.2	73.6	2142	21	AAC58369	Human PRO341 nucle
	20	16.2	73.6	2142	21	AAZ64947	Membrane-bound pro
	21	16.2	73.6	2142	22	AAF44093	Human PRO341 (UNQ3
	22	16.2	73.6	2150	22	AAF33221	Human secreted pro
c	23	16.2	73.6	63563	22	AAF28546	Genomic fragment #
c	24	15.8	71.8	552	21	AAA38400	Pseudomonas sp. WF
	25	15.8	71.8	1553	22	AAL37392	Human musculoskele
c	26	15.8	71.8	1880	13	AAQ24466	NANB hepatitis vir
	27	15.8	71.8	5822	20	AAV62933	Human mDia Rho tar
c	28	15.8	71.8	5822	20	AAV62934	Human mDia Rho tar
c	29	15.8	71.8	11279	21	AAA38389	Pseudomonas sp. WF
c	30	15.8	71.8	34185	21	AAC62130	Nucleotide sequenc
c	31	15.8	71.8	58857	21	AAA58471	Nucleotide sequenc
	32	15.6	70.9	28	22	AAF84061	5' and 3' targetin
	33	15.6	70.9	31	22	AAF84060	5' and 3' targetin
	34	15.6	70.9	211	20	AAH86896	Human single nucle
	35	15.6	70.9	211	20	AAH86897	Human single nucle
c	36	15.6	70.9	481	22	ABA09347	Human K channel TA
	37	15.6	70.9	567	22	ABA17873	Human nervous syst
	38	15.6	70.9	600	22	AAK75765	Human immune/haema
	39	15.6	70.9	600	22	AAK75766	Human immune/haema
	40	15.6	70.9	700	22	AAH92884	Human inflammatory
c	41	15.6	70.9	941	16	AAT02495	Pseudomonas lipase
	42	15.6	70.9	1025	23	AAS79585	DNA encoding novel
	43	15.6	70.9	1170	22	AAS00872	Monkey DNA encodin
	44	15.6	70.9	1215	21	AAA60774	Human SENR ligand
	45	15.6	70.9	1215	22	AAI64682	Human GPR14 encodi

ALIGNMENTS

RESULT 1
AAS01678/c
ID AAS01678 standard; DNA; 2093 BP.
XX
AC AAS01678;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human T-type calcium channel CACNA1G CpG island.
XX
KW Human; T-type calcium channel; CACNA1G; cytosine methylation; CpG island;
KW cellular proliferative disorder; colorectal cancer; age related disease;
KW apolipoprotein B; APOB; caudal type homeobox transcription factor 2;
KW CDX2; epidermal growth factor receptor; EGFR; fibrillin-1; FBN1;
KW G protein-coupled receptor 37; GPR37; heat shock 70kD protein 6; HSP70B';
KW HSPA6; RasGAP-related protein; IQGAP2; proteinase-activated receptor 2;
KW PAR2; paired-like homeodomain transcription factor 2; PITX2; klotho; KL;
KW patched A; patched B; PTCHA; PTCHB; syndecan 1; syndecan 4; SDC1; SDC4;
KW chromosome 17; ds.
XX
OS Homo sapiens.
XX
PN WO200119845-A1.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US25479.
XX
PR 15-SEP-1999; 99US-0398522.
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX

PI Issa J;
XX
DR WPI; 2001-244777/25.
XX
PT New nucleic acid molecule for use as a marker for screening cancer,
PT comprises the coding region for a T-type calcium channel and regulatory
PT sequences associated with the channel -
XX
PS Claim 32; Fig 4A; 125pp; English.
XX
CC The present sequence for human T-type calcium channel CACNA1G CpG island
CC is used to study the methylation state of the novel CACNA1G which maps to
CC chromosome 17. The methylation state of specific regions within CpG
CC islands associated with the CACNA1G gene correlate with several cancerous
CC phenotypes involving various tissue and cell types. Since aberrant
CC methylation of normally unmethylated CpG islands is often observed in
CC immortalised and transformed cells, CACNA1G is implicated in cellular
CC proliferative disorders e.g. leukaemia, colorectal, lung, breast and
CC other cancers. The nucleic acid coding for CACNA1G is useful as a marker
CC for screening cancer and age related diseases. A diagnostic kit
CC containing primers (AAS01574-AAS01623) for amplification of a
CC CpG-containing nucleic acid, where the primer hybridises with a target
CC polynucleotide sequence (AAS01627-AAS01676), can be used for detecting
CC aberrant methylation. The CpG island sequences (AAS01677-AAS01692) are
CC selected from genes encoding CACNA1G, apolipoprotein B (APOB), caudal
CC type homeobox transcription factor 2 (CDX2), epidermal growth factor
CC receptor (EGFR), fibrillin-1 (FBN1), G protein-coupled receptor 37
CC (GPR37), heat shock 70kD protein 6 (HSP70B'; HSPA6), RasGAP-related
CC protein (IQGAP2), klotho (KL), proteinase-activated receptor 2 (PAR2),
CC paired-like homeodomain transcription factor 2 (PITX2), patched A and B
CC (PTCHA; PTCHB) and syndecan 1 and 4 (SDC1; SDC4) or a MINT31 sequence.
XX
SQ Sequence 2093 BP; 295 A; 685 C; 813 G; 300 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 2093;
Best Local Similarity 86.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||||| ||| ||||||||
Db     728 GCAGCGGCGGAGGGCGGGGGTC 707
```

RESULT 2
AAL36954
ID AAL36954 standard; DNA; 5846 BP.
XX
AC AAL36954;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3319.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.

```
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451937/48.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the musculoskeletal system including
PT   musculoskeletal cancers and also for testing and detection e.g.
PT   diagnosis -
XX
PS   Example 2; SEQ ID NO 3319; 781pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC   (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment
CC   and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC   other cancers of the adrenal gland, bone, bone marrow, breast,
CC   gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5846 BP; 1109 A; 1894 C; 1739 G; 1104 T; 0 other;


  Query Match             78.2%;  Score 17.2;  DB 22;  Length 5846;
  Best Local Similarity   86.4%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||||| ||| ||||||||
Db     971 gcagcggcggagggcgggggtc 992


RESULT   3
AAL36955
ID   AAL36955 standard; DNA; 5848 BP.
XX
AC   AAL36955;
XX
DT   08-JAN-2002  (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 3320.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01338.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
```

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451937/48.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
XX
PS Example 2; SEQ ID NO 3320; 781pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;

```
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5848 BP; 1109 A; 1893 C; 1740 G; 1106 T; 0 other;


  Query Match             78.2%;  Score 17.2;  DB 22;  Length 5848;
  Best Local Similarity   86.4%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||||| ||| ||||||||
Db     972 gcagcggcggagggcgggggtc 993


RESULT   4
AAD04756/c
ID   AAD04756 standard; cDNA; 7741 BP.
XX
AC   AAD04756;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human T-type low voltage activated calcium channel alpha1G-c cDNA.
XX
KW   Human T-type low voltage activated calcium channel alpha1G-c;
KW   stress; epilepsy; schizophrenia; depression; sleep disorder;
KW   Cushing's disease; endocrine disorder; respiratory disorder;
KW   peripheral muscle disorder; muscle excitability; fertilisation;
KW   contraception; hypertension; neuronal firing regulation;
KW   cardiovascular disorder; gene therapy; forensic analysis;
KW   epidemiological study; neuroleptic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..522
FT                   /*tag= a
FT   CDS             523..7344
FT                   /*tag= b
FT                   /product= "Human T-type low voltage activated calcium
FT                   channel alpha1G-c protein"
FT                   /transl_except= (pos:3934..3936, aa:Gln)
FT                   /transl_except= (pos:3946..3948, aa:Gln)
FT                   /transl_except= (pos:5560..5562, aa:Pro)
FT                   /transl_except= (pos:5569..5571, aa:Gln)
FT                   /note= "The CDS is specifically claimed in
FT                   claim 2 as SEQ ID NO:3"
FT   3'UTR           7345..7741
FT                   /*tag= c
XX
PN   WO200130844-A1.
XX
PD   03-MAY-2001.
XX
PF   06-OCT-2000; 2000WO-US27761.
XX
PR   26-OCT-1999;   99US-0426998.
XX
PA   (ORTH ) ORTHO-MCNEIL PHARM INC.
XX
PI   Dubin AE,  Galindo JE,  Pyati J,  Zhu JY,  Erlander MG;
XX
DR   WPI; 2001-300486/31.
DR   P-PSDB; AAE01019.
XX
PT   New nucleic acid encoding human calcium channel protein, useful for
PT   identifying specific modulators and potential pharmaceuticals for
PT   treating e.g. epilepsy  -
XX
PS   Claim 2; Page 76-81; 115pp; English.
XX
CC   The invention relates to isoform of human T-type low voltage activated
CC   calcium channel (alpha1G-c) cDNA and protein. Cells transformed with
CC   calcium channel DNA to express calcium alpha1G-c channel protein are
CC   used to identify specific modulators (antagonists or agonists). These
CC   modulators are useful as therapeutic agents and are used for treating
CC   wide range of calcium alpha1G-c channel-mediated disorders, e.g. stress
CC   epilepsy, schizophrenia, depression, sleep disorders, Cushing's disease,
CC   endocrine disorders, respiratory disorder, peripheral muscle disorder,
CC   muscle excitability, fertilisation, contraception, disorders involving
CC   hypertension, neuronal firing regulation, potentiation of synaptic
CC   signals and cardiovascular disorders (e.g. atherosclerosis, cardiac
CC   hypertrophy, angina pectoris). Calcium alpha1G-c channel DNA is useful
CC   for isolating and identifying related molecule mutations. It is also
CC   optionally used as antisense sequences, in gene therapy. Calcium channel
CC   alpha1G-c DNA, protein and antibodies are useful for forensic analysis,
CC   diagnosis and epidemiological studies, by standard hybridisation or
CC   immunological assays.
CC   The present sequence is T-type low voltage activated calcium channel
CC   alpha1G-c cDNA. This sequence is isolated from human thalamus cDNA
CC   library.
XX
SQ   Sequence 7741 BP; 1469 A; 2496 C; 2287 G; 1489 T; 0 other;


  Query Match             78.2%;  Score 17.2;  DB 22;  Length 7741;
  Best Local Similarity   86.4%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||||| ||| ||||||||
Db     147 GCAGCGGCGGAGGGCGGGGGTC 126


RESULT   5
AAK67239/c
ID   AAK67239 standard; DNA; 30393 BP.
XX
AC   AAK67239;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
```

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,

```
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 other;


  Query Match             78.2%;  Score 17.2;  DB 22;  Length 30393;
  Best Local Similarity   86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           ||||||| | |||||| |||||
Db    8032 GAAGCGGTGGAGGCCGTGGGTC 8011


RESULT   6
AAS25846/c
ID   AAS25846 standard; cDNA; 664 BP.
XX
AC   AAS25846;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 25.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
```

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR P-PSDB; AAU15859.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; SEQ ID No 25; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

```
  Query Match           73.6%;  Score 16.2;  DB 22;  Length 664;
  Best Local Similarity  85.7%;  Pred. No. 6e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           ||||||||||||||| ||  ||
Db      49 AAGCGGCGTAGGCCAGGCCTC 29
```

RESULT 7
AAK92231
ID AAK92231 standard; cDNA; 780 BP.
XX
AC AAK92231;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 691.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens..
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 691; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 780 BP; 145 A; 240 C; 222 G; 168 T; 5 other;

```
  Query Match           73.6%;  Score 16.2;  DB 22;  Length 780;
  Best Local Similarity  85.7%;  Pred. No. 5.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           ||||||||| | ||| |||||
Db     129 aagcggcgtggcccgcgggtc 149
```

RESULT 8

```
AAH05059
ID   AAH05059 standard; cDNA; 786 BP.
XX
AC   AAH05059;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:1894.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 1894; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 786 BP; 136 A; 248 C; 255 G; 144 T; 3 other;


  Query Match             73.6%;  Score 16.2;  DB 22;  Length 786;
  Best Local Similarity   85.7%;  Pred. No. 5.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           ||||||||| | ||| |||||
Db     125 aagcggcgtggcccgcgggtc 145
```

```
RESULT   9
AAH05438
ID   AAH05438 standard; cDNA; 804 BP.
XX
AC   AAH05438;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:2273.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 2273; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 804 BP; 168 A; 246 C; 217 G; 169 T; 4 other;


  Query Match             73.6%;  Score 16.2;  DB 22;  Length 804;
  Best Local Similarity   85.7%;  Pred. No. 5.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           ||||||||| | ||| |||||
```

```
Db      66 aagcggcgtggcccgcgggtc 86


RESULT  10
AAK93381
ID   AAK93381 standard; cDNA; 865 BP.
XX
AC   AAK93381;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human cDNA clone representative sequence, SEQ ID NO: 1841.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Example 11; SEQ ID NO 1841; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence was used as the
CC   representative sequence from a human clone which was used in
CC   homology searches to identify the clone.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence 865 BP; 176 A; 263 C; 239 G; 182 T; 5 other;

  Query Match             73.6%;  Score 16.2;  DB 22;  Length 865;
  Best Local Similarity   85.7%;  Pred. No. 5.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           ||||||||| | ||| |||||
Db     123 aagcggcgtggcccgcgggtc 143


RESULT  11
ABA90359/c
ID   ABA90359 standard; cDNA; 1419 BP.
XX
AC   ABA90359;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Human polynucleotide #34.
XX
KW   Human; nootropic; neuroprotective; anticonvulsant; antidepressant;
KW   neuroleptic; tranquiliser; antiarrhythmic; cardiant; antiasthmatic;
KW   antiinflammatory; antilipaemic; hepatotropic; virucide; antidiabetic;
KW   nephrotropic; anorectic; cytostatic; vaccine; neurological disease;
KW   cardiovascular disease; respiratory disease; liver disease;
KW   renal disease; skeletal muscle disease; gastrointestinal disease;
KW   placental disease; testicular cancer; male fertility; pancreatic disease;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200181363-A1.
XX
PD   01-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13360.
XX
PR   27-APR-2000; 2000US-199963P.
PR   11-MAY-2000; 2000US-203336P.
PR   25-MAY-2000; 2000US-207087P.
PR   26-MAY-2000; 2000US-207546P.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Murdock PR,  Rizvi SK,  Smith RF,  Xiang Z,  Kabnick KS;
PI   Lai Y,  Xie Q;
XX
DR   WPI; 2002-041392/05.
DR   P-PSDB; ABB53294.
XX
PT   Novel polypeptides and polynucleotides useful as a vaccine for
PT   preventing and treating diseases associated the polypeptide, e.g.
PT   Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
PT   asthma, amnesias -
XX
PS   Claim 2; Page 63; 116pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a 277, 480,
CC   583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
CC   844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,
CC   784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
CC   given in the specification. The polypeptides, modulators of the
CC   polypeptides and antibodies against the polypeptides are useful for
CC   treating diseases such as neurological and psychiatric diseases
CC   including Alzheimer's, parasupranuclear palsy, Huntington's disease,
CC   myotonic dystrophy, anorexia and depression; cardiovascular diseases
CC   including congestive heart failure, Hodgkin's disease and myocardial
CC   infarction; respiratory diseases including asthma, chronic obstructive
CC   pulmonary disease, cystic fibrosis and adult respiratory distress
CC   syndrome; liver diseases including hypercholesterolaemia, cirrhosis,
CC   viral and nonviral hepatitis, Type II diabetes mellitis, and impaired
CC   glucose tolerance; renal disease including renal failure, acute tubular
CC   necrosis and glomerulonephritis; skeletal muscle diseases including
CC   Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal
CC   diseases including myotonia congenita and intestinal obstruction; lymph
CC   diseases including lymphagiectasia; diseases of placenta including
CC   choriocarcinoma; diseases of testes including testicular cancer,
CC   male reproductive diseases including low testosterone and male
CC   infertility; and disease of pancreas including diabetic ketoacidosis,
CC   Type 1 and 2 diabetes and obesity. The present sequence encodes a
CC   polypeptide of the invention.
XX
SQ   Sequence 1419 BP; 261 A; 480 C; 447 G; 231 T; 0 other;

  Query Match             73.6%;  Score 16.2;  DB 24;  Length 1419;
  Best Local Similarity   85.7%;  Pred. No. 5.6e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
```

```
                ||||||||||||||| ||  ||
Db    1129 AAGCGGCGTAGGCCAGGCCTC 1109


RESULT  12
AAH15691
ID   AAH15691 standard; cDNA; 1768 BP.
XX
AC   AAH15691;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:14060.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14060; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1768 BP; 351 A; 487 C; 571 G; 359 T; 0 other;


  Query Match             73.6%;  Score 16.2;  DB 22;  Length 1768;
  Best Local Similarity   85.7%;  Pred. No. 5.5e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           ||||||||| | ||| |||||
Db     125 aagcggcgtggcccgcgggtc 145


RESULT  13
ABA90358/c
ID   ABA90358 standard; cDNA; 1782 BP.
XX
AC   ABA90358;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Human polynucleotide #33.
XX
KW   Human; nootropic; neuroprotective; anticonvulsant; antidepressant;
KW   neuroleptic; tranquiliser; antiarrhythmic; cardiant; antiasthmatic;
KW   antiinflammatory; antilipaemic; hepatotropic; virucide; antidiabetic;
KW   nephrotropic; anorectic; cytostatic; vaccine; neurological disease;
KW   cardiovascular disease; respiratory disease; liver disease;
KW   renal disease; skeletal muscle disease; gastrointestinal disease;
KW   placental disease; testicular cancer; male fertility; pancreatic disease;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200181363-A1.
XX
PD   01-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13360.
XX
PR   27-APR-2000; 2000US-199963P.
PR   11-MAY-2000; 2000US-203336P.
PR   25-MAY-2000; 2000US-207087P.
PR   26-MAY-2000; 2000US-207546P.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Murdock PR,  Rizvi SK,  Smith RF,  Xiang Z,  Kabnick KS;
PI   Lai Y,  Xie Q;
XX
DR   WPI; 2002-041392/05.
DR   P-PSDB; ABB53293.
XX
PT   Novel polypeptides and polynucleotides useful as a vaccine for
PT   preventing and treating diseases associated the polypeptide, e.g.
PT   Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
PT   asthma, amnesias -
XX
PS   Claim 2; Page 62-63; 116pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a 277, 480,
CC   583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
CC   844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,
CC   784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
CC   given in the specification. The polypeptides, modulators of the
CC   polypeptides and antibodies against the polypeptides are useful for
CC   treating diseases such as neurological and psychiatric diseases
CC   including Alzheimer's, parasupranuclear palsy, Huntington's disease,
CC   myotonic dystrophy, anorexia and depression; cardiovascular diseases
CC   including congestive heart failure, Hodgkin's disease and myocardial
CC   infarction; respiratory diseases including asthma, chronic obstructive
CC   pulmonary disease, cystic fibrosis and adult respiratory distress
CC   syndrome; liver diseases including hypercholesterolaemia, cirrhosis,
CC   viral and nonviral hepatitis, Type II diabetes mellitis, and impaired
CC   glucose tolerance; renal disease including renal failure, acute tubular
CC   necrosis and glomerulonephritis; skeletal muscle diseases including
CC   Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal
CC   diseases including myotonia congenita and intestinal obstruction; lymph
CC   diseases including lymphagiectasia; diseases of placenta including
```

```
CC   choriocarcinoma; diseases of testes including testicular cancer,
CC   male reproductive diseases including low testosterone and male
CC   infertility; and disease of pancreas including diabetic ketoacidosis,
CC   Type 1 and 2 diabetes and obesity. The present sequence encodes a
CC   polypeptide of the invention.
XX
SQ   Sequence 1782 BP; 311 A; 591 C; 569 G; 311 T; 0 other;

  Query Match             73.6%;  Score 16.2;  DB 24;  Length 1782;
  Best Local Similarity   85.7%;  Pred. No. 5.5e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 aagcggcgtaggccgggggtc 22
           |||||||||||||| ||  ||
Db    1492 AAGCGGCGTAGGCCAGGCCTC 1472


RESULT  14
ABL34375
ID   ABL34375 standard; DNA; 1790 BP.
XX
AC   ABL34375;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2348.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 2348; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 1790 BP; 232 A; 234 C; 676 G; 648 T; 0 other;

  Query Match             73.6%;  Score 16.2;  DB 24;  Length 1790;
  Best Local Similarity   85.7%;  Pred. No. 5.4e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggt 21
           |||| |||| ||| |||||||
Db     274 gaaggggcggagggcgggggt 294


RESULT  15
AAH15850
ID   AAH15850 standard; cDNA; 1993 BP.
XX
AC   AAH15850;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:14353.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14353; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1993 BP; 408 A; 594 C; 525 G; 466 T; 0 other;
```

Query Match 73.6%; Score 16.2; DB 22; Length 1993;
Best Local Similarity 85.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      2 aagcggcgtaggccgggggtc 22
          ||||||||| | ||| |||||
Db     66 aagcggcgtggcccgcgggtc 86
```

Search completed: June 23, 2002, 15:01:16
Job time: 81208 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 23, 2002, 15:09:16 ; Search time 167.81 Seconds
(without alignments)
32.203 Million cell updates/sec

Title: US-09-747-514A-4
Perfect score: 22
Sequence: 1 gaagcggcgtaggccgggggtc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	17.2	78.2	7741	4	US-09-426-998-4	Sequence 4, Appli
	2	15.8	71.8	5822	3	US-08-899-595-4	Sequence 4, Appli
c	3	15.8	71.8	5822	3	US-08-899-595-5	Sequence 5, Appli
c	4	15.6	70.9	942	3	US-08-732-412-1	Sequence 1, Appli
	5	15.6	70.9	2126	2	US-08-789-354-1	Sequence 1, Appli
	6	15.6	70.9	2126	3	US-09-110-937-1	Sequence 1, Appli
	7	15.6	70.9	2126	3	US-09-058-725B-1	Sequence 1, Appli
	8	15.6	70.9	2126	3	US-09-232-857-1	Sequence 1, Appli
c	9	15.6	70.9	3546	4	US-08-872-757-3	Sequence 3, Appli
c	10	15.2	69.1	1247	5	PCT-US91-02766-19	Sequence 19, Appl
c	11	15.2	69.1	1247	6	5219759-3	Patent No. 5219759
c	12	15.2	69.1	1316	5	PCT-US91-02766-21	Sequence 21, Appl
c	13	15.2	69.1	1316	6	5219759-1	Patent No. 5219759
	14	15.2	69.1	4378	2	US-09-080-897-3	Sequence 3, Appli
	15	15.2	69.1	4378	4	US-09-323-735-3	Sequence 3, Appli
	16	15.2	69.1	4399	3	US-08-899-595-2	Sequence 2, Appli
c	17	15.2	69.1	6854	4	US-09-194-905-7	Sequence 7, Appli
c	18	15.2	69.1	12412	1	US-08-390-878-18	Sequence 18, Appl
	19	15.2	69.1	22108	3	US-09-053-197A-3	Sequence 3, Appli
	20	15.2	69.1	22108	4	US-09-085-761A-3	Sequence 3, Appli
c	21	15.2	69.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	22	14.8	67.3	717	4	US-09-124-238A-7	Sequence 7, Appli
	23	14.8	67.3	1620	4	US-09-124-238A-32	Sequence 32, Appl
	24	14.8	67.3	1644	4	US-09-124-238A-9	Sequence 9, Appli
	25	14.8	67.3	1665	4	US-09-124-238A-33	Sequence 33, Appl
	26	14.8	67.3	1689	4	US-09-124-238A-22	Sequence 22, Appl
	27	14.8	67.3	4208	4	US-09-214-278-6	Sequence 6, Appli
	28	14.8	67.3	4208	4	US-09-068-740A-10	Sequence 10, Appl
	29	14.8	67.3	5590	3	US-08-882-046-1	Sequence 1, Appli
	30	14.8	67.3	6464	2	US-08-400-159-5	Sequence 5, Appli
	31	14.8	67.3	6464	3	US-08-611-729A-5	Sequence 5, Appli
	32	14.6	66.4	63	2	US-08-284-391B-51	Sequence 51, Appl
	33	14.6	66.4	63	4	US-09-218-950-51	Sequence 51, Appl
c	34	14.6	66.4	387	4	US-08-943-136-1	Sequence 1, Appli
c	35	14.6	66.4	387	4	US-08-973-518-1	Sequence 1, Appli
c	36	14.6	66.4	462	3	US-08-854-531-1	Sequence 1, Appli
c	37	14.6	66.4	462	5	PCT-US95-13552-1	Sequence 1, Appli
c	38	14.6	66.4	549	3	US-08-441-971-58	Sequence 58, Appl
c	39	14.6	66.4	549	3	US-08-441-971-59	Sequence 59, Appl
c	40	14.6	66.4	549	3	US-08-441-971-60	Sequence 60, Appl
c	41	14.6	66.4	549	3	US-08-441-971-61	Sequence 61, Appl
c	42	14.6	66.4	549	3	US-08-441-971-64	Sequence 64, Appl
c	43	14.6	66.4	549	4	US-08-221-653-58	Sequence 58, Appl
c	44	14.6	66.4	549	4	US-08-221-653-59	Sequence 59, Appl
c	45	14.6	66.4	549	4	US-08-221-653-60	Sequence 60, Appl

ALIGNMENTS

```
RESULT   1
US-09-426-998-4/c
; Sequence 4, Application US/09426998
; Patent No. 6358706
; GENERAL INFORMATION:
;  APPLICANT: DUBIN, ADRIENNE E.
;  APPLICANT:  PYATI, JAYASHREE
;  APPLICANT:  ZHU, JESSICA Y
;  APPLICANT:  ERLANDER, MARK G
;  APPLICANT:  GALINDO, JOSE E
;  TITLE OF INVENTION: DNA ENCODING HUMAN ALPHA1G T-TYPE CALCIUM
;  TITLE OF INVENTION:  CHANNEL (ALPHA1G-C)
;  FILE REFERENCE: ORT-1057
;  CURRENT APPLICATION NUMBER: US/09/426,998
;  CURRENT FILING DATE:  1999-10-26
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PATENTIN VER. 2.0
; SEQ ID NO 4
;   LENGTH: 7741
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-426-998-4


  Query Match             78.2%;  Score 17.2;  DB 4;  Length 7741;
  Best Local Similarity   86.4%;  Pred. No. 31;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||||| ||| ||||||||
Db     147 GCAGCGGCGGAGGGCGGGGGTC 126


RESULT   2
US-08-899-595-4
; Sequence 4, Application US/08899595
; Patent No. 6111072
;  GENERAL INFORMATION:
;    APPLICANT:  Narumiya, Shuh
;    APPLICANT:  Takahashi, No. 6111072uaki
;    TITLE OF INVENTION:  RHO TARGET PROTEIN HUMAN MDIA AND GENE
;    TITLE OF INVENTION:  ENCODING SAME
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,595
;      FILING DATE:  24-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-242701
;      FILING DATE:  26-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 9-90170
;      FILING DATE:  25-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stephen A. Bent
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  049441/0112
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Human
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..3972
US-08-899-595-4


  Query Match            71.8%;  Score 15.8;  DB 3;  Length 5822;
  Best Local Similarity  89.5%;  Pred. No. 1.2e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 gcggcgtaggccgggggtc 22
           ||||||||||| ||||||
Db     143 GCGGCGTAGGCGCGGGGTC 161


RESULT   3
US-08-899-595-5/c
; Sequence 5, Application US/08899595
; Patent No. 6111072
;  GENERAL INFORMATION:
;    APPLICANT:  Narumiya, Shuh
;    APPLICANT:  Takahashi, No. 6111072uaki
;    TITLE OF INVENTION:  RHO TARGET PROTEIN HUMAN MDIA AND GENE
;    TITLE OF INVENTION:  ENCODING SAME
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,595
;      FILING DATE:  24-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-242701
;      FILING DATE:  26-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 9-90170
;      FILING DATE:  25-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stephen A. Bent
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  049441/0112
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ANTI-SENSE:  YES
;    ORIGINAL SOURCE:
;      ORGANISM:  Human
US-08-899-595-5


  Query Match            71.8%;  Score 15.8;  DB 3;  Length 5822;
  Best Local Similarity  89.5%;  Pred. No. 1.2e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 gcggcgtaggccgggggtc 22
           ||||||||||| ||||||
Db    5680 GCGGCGTAGGCGCGGGGTC 5662


RESULT   4
US-08-732-412-1/c
; Sequence 1, Application US/08732412
; Patent No. 6017866
;  GENERAL INFORMATION:
;    APPLICANT:  Aehle, Wolfgang
;    APPLICANT:  Gerritse, Gijsbert
;    APPLICANT:  Lenting, Hermanus
;    TITLE OF INVENTION:  LIPASES WITH IMPROVED
;    TITLE OF INVENTION:  SURFACTANT RESISTANCE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genencor International, Inc.
;      STREET:  925 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1013
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/732,412
;      FILING DATE:  22-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP95/01687
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 94201253.5
;      FILING DATE:  04-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stone, Christopher L.
;      REGISTRATION NUMBER:  35,696
;      REFERENCE/DOCKET NUMBER:  GC307-US
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  650-846-7620
;       TELEFAX:  650-845-6504
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  942 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-732-412-1


  Query Match             70.9%;  Score 15.6;  DB 3;  Length 942;
  Best Local Similarity   81.8%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           ||||| | | ||||||| |||||
Db     144 GAAGCCGAGCAGGCCGTGGGTC 123


RESULT   5
US-08-789-354-1
; Sequence 1, Application US/08789354
; Patent No. 5851798
;  GENERAL INFORMATION:
;    APPLICANT:  Shabon, Usman
;    APPLICANT:  Bergsma, Derk
;    TITLE OF INVENTION:  Cloning of Human GPR14 Re
;    TITLE OF INVENTION:  ceptor
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/789,354
;      FILING DATE:  27-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  P50610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-4026
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-789-354-1


  Query Match             70.9%;  Score 15.6;  DB 2;  Length 2126;
  Best Local Similarity   81.8%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||| | |||||||||| |

Db    1694 GCAGCGGGGGAGGCCGGGGGCC 1715


RESULT   6
US-09-110-937-1
; Sequence 1, Application US/09110937A
; Patent No. 6005074
; GENERAL INFORMATION:
;  APPLICANT: SHABON, USMAN
;  APPLICANT: BERGSMA, DERK
;  TITLE OF INVENTION: CLONING OF HUMAN GPR14 RECEPTOR
;  FILE REFERENCE: P50610-1
;  CURRENT APPLICATION NUMBER: US/09/110,937A
;  CURRENT FILING DATE: 1998-07-06
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2126
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-110-937-1


  Query Match             70.9%;  Score 15.6;  DB 3;  Length 2126;
  Best Local Similarity   81.8%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccgggggtc 22
           | ||||| | |||||||||| |
Db    1694 gcagcgggggaggccgggggcc 1715


RESULT   7
US-09-058-725B-1
; Sequence 1, Application US/09058725B
; Patent No. 6133420
;  GENERAL INFORMATION:
;    APPLICANT:  Ames, Robert
;    APPLICANT:  Sarau, Henry
;    APPLICANT:  Foley, James
;    APPLICANT:  Chamber, Jon
;    TITLE OF INVENTION:  A Method of Finding Angonist
;    TITLE OF INVENTION:  and Antagonist To Human and Rat GPR 14
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/058,725B
;      FILING DATE:  April 10, 1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/789,354
;      FILING DATE:  27-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T.
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  GP50005-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

; LENGTH: 2126 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-058-725B-1

Query Match 70.9%; Score 15.6; DB 3; Length 2126;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaagcggcgtaggccgggggtc 22
| ||||| | |||||||||| |
Db 1694 GCAGCGGGGGAGGCCGGGGGCC 1715

RESULT 8
US-09-232-857-1
; Sequence 1, Application US/09232857
; Patent No. 6159700
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS, STEPHEN
; APPLICANT: WILLETTE, ROBERT
; APPLICANT: AIYAR, NAMBI
; APPLICANT: ROMANIC, ANNE
; APPLICANT: KHANDOUDI, NASSIRAH
; APPLICANT: GOUT, BERNARD
; APPLICANT: AL-BARAZANJI, KAMAL
; APPLICANT: AMES, ROBERT S.
; APPLICANT: FOLEY, JAMES J.
; APPLICANT: SARAU, HENRY
; APPLICANT: CHAMBERS, JON K.
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMA, DERK
; TITLE OF INVENTION: A METHOD OF FINDING AGONIST
; TITLE OF INVENTION: AND ANTAGONIST TO HUMAN AND RAT GPR14
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/232,857
; FILING DATE: 15-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/789,354
; FILING DATE: 27-JAN-1997
; APPLICATION NUMBER: 60/074,075
; FILING DATE: 09-FEB-1998
; APPLICATION NUMBER: 09/058,725
; FILING DATE: 10-APR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-50005-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2126 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-232-857-1

Query Match 70.9%; Score 15.6; DB 3; Length 2126;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaagcggcgtaggccgggggtc 22
| ||||| | |||||||||| |
Db 1694 GCAGCGGGGGAGGCCGGGGGCC 1715

RESULT 9
US-08-872-757-3/c
; Sequence 3, Application US/08872757
; Patent No. 6258584
; GENERAL INFORMATION:
; APPLICANT: Prockop, Darwin J.
; APPLICANT: Hojima, Yoshio
; APPLICANT: Li, Shi-Wu
; APPLICANT: Sieron, Aleksander
; TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
; TITLE OF INVENTION: PROCESSES; METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/872,757
; FILING DATE: 10-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,187
; FILING DATE: 01-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8389-028-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3546 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2958
US-08-872-757-3

Query Match 70.9%; Score 15.6; DB 4; Length 3546;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaagcggcgtaggccgggggtc 22
|||||||| ||| |||||| |

```
Db    2463 GAAGCGGCCGAGGACGGGGGCC 2442


RESULT  10
PCT-US91-02766-19/c
; Sequence 19, Application PC/TUS9102766
;  GENERAL INFORMATION:
;     APPLICANT:  NASCIMENTO, CARLOS G.
;     APPLICANT:  CALDERON-CACIA, MARIA D.
;     TITLE OF INVENTION:  GLYCOSYLATED PDGF
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Irell & Manella
;       STREET:  545 Middlefield Road, Suite 200
;       CITY:  Menlo Park
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94025
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/02766
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/515,474
;       FILING DATE:  26-APR-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  ROBINS, ROBERTA L.
;       REGISTRATION NUMBER:  33,208
;       REFERENCE/DOCKET NUMBER:  2300-0105.40
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 327-7250
;       TELEFAX:  (415) 327-2951
;       TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1247 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  396..983
PCT-US91-02766-19


  Query Match             69.1%;  Score 15.2;  DB 5;  Length 1247;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccggggg 20
           |||| |||| ||| ||||||
Db     131 GAAGGGGCGGAGGGCGGGGG 112


RESULT  11
5219759-3/c
;Patent No. 5219759
;     APPLICANT: HELDIN, CARL-HENRIK;BETSHOLTZ, CHRISTER;WESTERMARK,
;BENGT;KNOTT, TIMOTHY J.;SCOTT, JAMES;BELL, GRAEME J.
;     TITLE OF INVENTION: RECOMBINANT DNA ENCODING PDGF A-CHAIN
;POLYPEPTIDE AND EXPRESSION VECTORS
;     NUMBER OF SEQUENCES: 18
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/574,540
;       FILING DATE: 27-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 41,299
;       FILING DATE: 22-APR-1987
;SEQ ID NO:3:
;       LENGTH: 1247
5219759-3


  Query Match             69.1%;  Score 15.2;  DB 6;  Length 1247;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccggggg 20
           |||| |||| ||| ||||||
Db     131 GAAGGGGCGGAGGGCGGGGG 112


RESULT  12
PCT-US91-02766-21/c
; Sequence 21, Application PC/TUS9102766
;  GENERAL INFORMATION:
;     APPLICANT:  NASCIMENTO, CARLOS G.
;     APPLICANT:  CALDERON-CACIA, MARIA D.
;     TITLE OF INVENTION:  GLYCOSYLATED PDGF
;     NUMBER OF SEQUENCES:  24
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Irell & Manella
;       STREET:  545 Middlefield Road, Suite 200
;       CITY:  Menlo Park
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94025
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/02766
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/515,474
;       FILING DATE:  26-APR-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  ROBINS, ROBERTA L.
;       REGISTRATION NUMBER:  33,208
;       REFERENCE/DOCKET NUMBER:  2300-0105.40
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 327-7250
;       TELEFAX:  (415) 327-2951
;       TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1316 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  396..1028
PCT-US91-02766-21


  Query Match             69.1%;  Score 15.2;  DB 5;  Length 1316;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccggggg 20
           |||| |||| ||| ||||||
Db     131 GAAGGGGCGGAGGGCGGGGG 112


RESULT  13
5219759-1/c
;Patent No. 5219759
```

```
;    APPLICANT: HELDIN, CARL-HENRIK;BETSHOLTZ, CHRISTER;WESTERMARK,
;BENGT;KNOTT, TIMOTHY J.;SCOTT, JAMES;BELL, GRAEME J.
;    TITLE OF INVENTION: RECOMBINANT DNA ENCODING PDGF A-CHAIN
;POLYPEPTIDE AND EXPRESSION VECTORS
;    NUMBER OF SEQUENCES: 18
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/574,540
;      FILING DATE: 27-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 41,299
;      FILING DATE: 22-APR-1987
;SEQ ID NO:1:
;      LENGTH: 1316
5219759-1


  Query Match             69.1%;  Score 15.2;  DB 6;  Length 1316;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gaagcggcgtaggccggggg 20
           |||| |||| ||| ||||||
Db     131 GAAGGGGCGGAGGGCGGGGG 112


RESULT  14
US-09-080-897-3
; Sequence 3, Application US/09080897
; Patent No. 5985574
;  GENERAL INFORMATION:
;    APPLICANT:  King, Mary-Claire
;    APPLICANT:  Lynch, Eric D.
;    APPLICANT:  Lee, Ming
;    APPLICANT:  Morrow, Jan E.
;    APPLICANT:  Welcsh, Piri L.
;    APPLICANT:  Leon, Pedro E.
;    TITLE OF INVENTION:  Modulators of Actin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/080,897
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UW97-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-080-897-3


  Query Match             69.1%;  Score 15.2;  DB 2;  Length 4378;
  Best Local Similarity   85.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 agcggcgtaggccgggggtc 22
           |||||||||| | ||||||
Db      52 AGCGGCGTAGACAAGGGGTC 71


RESULT  15
US-09-323-735-3
; Sequence 3, Application US/09323735
; Patent No. 6197932
;  GENERAL INFORMATION:
;    APPLICANT:  King, Mary-Claire
;    APPLICANT:  Lynch, Eric D.
;    APPLICANT:  Lee, Ming
;    APPLICANT:  Morrow, Jan E.
;    APPLICANT:  Welcsh, Piri L.
;    APPLICANT:  Leon, Pedro E.
;    TITLE OF INVENTION:  Modulators of Actin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/323,735
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/080,897
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UW97-001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-323-735-3


  Query Match             69.1%;  Score 15.2;  DB 4;  Length 4378;
  Best Local Similarity   85.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 agcggcgtaggccgggggtc 22
           |||||||||| | ||||||
Db      52 AGCGGCGTAGACAAGGGGTC 71


Search completed: June 23, 2002, 15:09:20
Job time: 78584 sec
```

EM 200104
ED Entered STN: 20010425
Last Updated on STN: 20010425
Entered Medline: 20010419

L11 ANSWER 5 OF 16 MEDLINE
AN 2001015399 MEDLINE
DN 20465132 PubMed ID: 11008205
TI Urokinase-type plasminogen activator and its receptor in colorectal cancer: independent prognostic factors of metastasis and cancer-specific survival and potential therapeutic targets.
AU Yang J L; Seetoo D q; Wang Y; Ranson M; Berney C R; Ham J M; Russell P J; Crowe P J
CS Department of Surgery, Faculty of Medicine of University of New South Wales, Prince of Wales Hospital, Randwick, Australia.. j.yang@unsw.edu.au
SO INTERNATIONAL JOURNAL OF CANCER, (2000 Sep 20) 89 (5) 431-9.
Journal code: 0042124. ISSN: 0020-7136.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200011
ED Entered STN: 20010322
Last Updated on STN: 20010322
Entered Medline: 20001101

L11 ANSWER 6 OF 16 MEDLINE
AN 1999290668 MEDLINE
DN 99290668 PubMed ID: 10364000
TI Prostaglandin J2 and 15-deoxy-delta12,14-prostaglandin J2 induce proliferation of cyclooxygenase-depleted colorectal cancer cells.
AU Chinery R; Coffey R J; Graves-Deal R; Kirkland S C; Sanchez S C; Zackert W E; Oates J A; Morrow J D
CS Department of Medicine, and The Vanderbilt Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA.
NC CA77839 (NCI)
DK48831 (NIDDK)
GM15431 (NIGMS)
+
SO CANCER RESEARCH, (1999 Jun 1) 59 (11) 2739-46.
Journal code: 2984705R. ISSN: 0008-5472.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199906
ED Entered STN: 19990714
Last Updated on STN: 19990714
Entered Medline: 19990629

L11 ANSWER 7 OF 16 MEDLINE
AN 93388005 MEDLINE
DN 93388005 PubMed ID: 8104163
TI UICC/**CRC** conference on targeted cancer therapy. Royal Free Hospital School of Medicine, London, December 17-19, 1991.
AU Begent R H
SO INTERNATIONAL JOURNAL OF CANCER, (1993 Sep 30) 55 (3) 355-8.
Journal code: 0042124. ISSN: 0020-7136.
CY United States
DT Conference; Conference Article; (CONGRESSES)
LA English
FS Priority Journals
EM 199310
ED Entered STN: 19931105

Last Updated on STN: 19990129
Entered Medline: 19931021

L11 ANSWER 8 OF 16 MEDLINE
AN 93285101 MEDLINE
DN 93285101 PubMed ID: 8389694
TI Linkage of superantigen-like stimulation of syngeneic T cells in a mouse model of follicular center B cell lymphoma to transcription of endogenous mammary tumor virus.
AU Tsiagbe V K; Yoshimoto T; Asakawa J; Cho S Y; Meruelo D; Thorbecke G J
CS Department of Pathology, Kaplan Comprehensive Cancer Center, New York University School of Medicine, NY 10016.
NC CA-14462 (NCI)
CA-22247 (NCI)
CA-31346 (NCI)
SO EMBO JOURNAL, (1993 Jun) 12 (6) 2313-20.
Journal code: 8208664. ISSN: 0261-4189.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
OS GENBANK-L11933
EM 199307
ED Entered STN: 19930723
Last Updated on STN: 19970203
Entered Medline: 19930715

L11 ANSWER 9 OF 16 MEDLINE
AN 93171257 MEDLINE
DN 93171257 PubMed ID: 8382205
TI Inositol 1,4,5-trisphosphate receptor expression in cardiac myocytes.
AU Moschella M C; Marks A R
CS Department of Medicine, Mount Sinai School of Medicine, New York 10029.
SO JOURNAL OF CELL BIOLOGY, (1993 Mar) 120 (5) 1137-46.
Journal code: 0375356. ISSN: 0021-9525.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199303
ED Entered STN: 19930402
Last Updated on STN: 19980206
Entered Medline: 19930323

L11 ANSWER 10 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 2001:833564 CAPLUS
DN 135:367649
TI Identification of sequence motifs in oligonucleotides with **antisense** activity by correlation analysis
IN Freier, Susan M.; Matveeva, Olga; Tsodikov, Alexander; Giddings, Michael C.; Wyatt, Jacqueline R.
PA Isis Pharmaceuticals, Inc., USA; University of Utah Research Foundation
SO PCT Int. Appl., 57 pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI WO 2001085996	A1	20011115	WO 2001-US14157	20010502

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO,

RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG
PRAI US 2000-568165 A 20000509
RE.CNT 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L11 ANSWER 11 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 2000:772742 CAPLUS
DN 133:330554
TI Protein and cDNA sequences of corn NPR1 gene and uses thereof in plant disease control
IN Crane, Edmund H., III; Rice, Douglas A.; Simmons, Carl R.; Tossberg, John T.; Sandahl, Gary A.; Zhang, Lingyu
PA Pioneer Hi-Bred International, Inc., USA
SO PCT Int. Appl., 86 pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 2000065037	A2	20001102	WO 2000-US10479	20000419
	WO 2000065037	A3	20010726		
	W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
	RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
	EP 1173575	A2	20020123	EP 2000-928204	20000419
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
	BR 2000009980	A	20020409	BR 2000-9980	20000419
PRAI	US 1999-130692P	P	19990423		
	WO 2000-US10479	W	20000419		

L11 ANSWER 12 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 2000:751508 CAPLUS
DN 134:205764
TI Urokinase-type plasminogen activator and its receptor in colorectal cancer: independent prognostic factors of metastasis and cancer-specific survival and potential therapeutic targets
AU Yang, Jia-Lin; Seetoo, Da-qiang; Wang, Yao; Ranson, Marie; Berney, Christophe R.; Ham, John M.; Russell, Pamela J.; Crowe, Philip J.
CS Department of Surgery, Faculty of Medicine of University of New South Wales, Randwick, Australia
SO International Journal of Cancer (2000), 89(5), 431-439
CODEN: IJCNAW; ISSN: 0020-7136
PB Wiley-Liss, Inc.
DT Journal
LA English
RE.CNT 52 THERE ARE 52 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L11 ANSWER 13 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 1999:405087 CAPLUS
DN 131:57413
TI Protein 14-3-3.sigma. arrest of the cell cycle provides the basis for

diagnostic assays and therapeutic compositions
IN Hermeking, Heiko; Vogelstein, Bert; Kinzler, Kenneth W.
PA The Johns Hopkins Univ., USA
SO PCT Int. Appl., 73 pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI WO 9931240	A2	19990624	WO 1998-US26924	19981218
WO 9931240	A3	19990902		
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
US 6335156	B1	20020101	US 1998-210748	19981215
CA 2315279	AA	19990624	CA 1998-2315279	19981218
AU 9918314	A1	19990705	AU 1999-18314	19981218
AU 744193	B2	20020221		
EP 1037987	A2	20000927	EP 1998-963256	19981218
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
PRAI US 1997-69416P	P	19971218		
US 1998-210748	A	19981215		
WO 1998-US26924	W	19981218		

L11 ANSWER 14 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 1999:368555 CAPLUS
DN 131:125831
TI Prostaglandin J2 and 15-deoxy-.DELTA.12,14-prostaglandin J2 induce proliferation of cyclooxygenase-depleted colorectal cancer cells
AU Chinery, Rebecca; Coffey, Robert J.; Graves-Deal, Ramona; Kirkland, Susan C.; Sanchez, Stephanie C.; Zackert, William E.; Oates, John A.; Morrow, Jason D.
CS Departments of Medicine and Cell Biology and The Vanderbilt Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, 37232, USA
SO Cancer Research (1999), 59(11), 2739-2746
CODEN: CNREA8; ISSN: 0008-5472
PB AACR Subscription Office
DT Journal
LA English
RE.CNT 36 THERE ARE 36 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L11 ANSWER 15 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 1999:92323 CAPLUS
TI Hydrolysis of GpppG, a model for the 5'-cap of mRNA by dinuclear metal ion complexes
AU Morrow, Janet R.; McCue, Kevin P.
CS Department of Chemistry, State University of New York, Amherst, NY, 14260-3000, USA
SO Book of Abstracts, 217th ACS National Meeting, Anaheim, Calif., March 21-25 (1999), INOR-506 Publisher: American Chemical Society, Washington, D. C.
CODEN: 67GHA6
DT Conference; Meeting Abstract
LA English

L11 ANSWER 16 OF 16 CAPLUS COPYRIGHT 2002 ACS
AN 1993:536993 CAPLUS
DN 119:136993
TI Linkage of superantigen-like stimulation of syngeneic T cells in a mouse model of follicular center B cell lymphoma to transcription of endogenous mammary tumor virus
AU Tsiagbe, V. K.; Yoshimoto, T.; Asakawa, J.; Cho, S. Y.; Meruelo, D.; Thorbecke, G. J.
CS Sch. Med., New York Univ., New York, NY, 10016, USA
SO EMBO J. (1993), 12(6), 2313-20
CODEN: EMJODG; ISSN: 0261-4189
DT Journal
LA English

=> d his

(FILE 'HOME' ENTERED AT 15:54:14 ON 25 JUN 2002)

FILE 'BIOSIS, MEDLINE, CAPLUS' ENTERED AT 15:55:44 ON 25 JUN 2002

L1	183033	PSEUDOMONAS
L2	78120	VIRULENCE
L3	3966	L1 AND L2
L4	2	L3 AND CRC
L5	93	CATABOLITE (W) REPRESSION (W) CONTROL
L6	1	L1 AND L2 AND L5
L7	4292	CRC
L8	2	L3 AND L7
L9	58738	ANTISENS?
L10	11697	RIBOZYM?
L11	16	L7 AND L9
L12	0	L7 AND L10